

RESEARCH ARTICLE

Sample size reestimation for clinical trials with longitudinal negative binomial counts including time trends

Thomas Asendorf¹ | Robin Henderson² | Heinz Schmidli³ | Tim Friede¹

¹Department of Medical Statistics,
University Medical Center Göttingen,
Göttingen, Germany

²School of Mathematics and Statistics,
University of Newcastle, Newcastle, United
Kingdom

³Statistical Methodology, Novartis Pharma
AG, Basel, Switzerland

Correspondence

Thomas Asendorf, Department of Medical
Statistics, Humboldtallee 32, 37073
Göttingen, Germany. Email:
thomas.asendorf@med.uni-goettingen.de

Present Address

Department of Medical Statistics,
Humboldtallee 32, 37073 Göttingen,
Germany.

Abstract

In some diseases, such as multiple sclerosis, lesion counts obtained from magnetic resonance imaging (MRI) are used as markers of disease progression. This leads to longitudinal, and typically overdispersed, count data outcomes in clinical trials. Models for such data invariably include a number of nuisance parameters which can be difficult to specify at the planning stage, leading to considerable uncertainty in sample size specification. Consequently, blinded sample size reestimation procedures are used, allowing for an adjustment of the sample size within an ongoing trial by estimating relevant nuisance parameters at an interim point, without compromising trial integrity. To date, the methods available for reestimation have required an assumption that the mean count is time-constant within patients. We propose a new modelling approach that maintains advantages of established procedures but allows for general underlying and treatment-specific time trends in the mean response. A simulation study is conducted to assess the effectiveness of blinded sample size reestimation methods over fixed designs. Sample sizes attained through blinded sample size reestimation procedures are shown to maintain the desired study power without inflating the type I error rate and the procedure is demonstrated on MRI data from a recent study in multiple sclerosis.

KEYWORDS:

adaptive design; lesion counts; sample size reestimation; negative binomial; gamma frailty

1 | INTRODUCTION

Endpoints attained through magnetic resonance imaging (MRI) are established outcome measures in early phase clinical trials.¹ For example, counts of T2 hyperintense lesions or gadolinium-enhanced lesions are accepted measures of inflammatory disease activity in multiple sclerosis (MS).² Lesion counts are often observed over regular time intervals, e.g. monthly scans, over the

⁰**Abbreviations:** MRI, magnetic resonance imaging; MS, multiple sclerosis; BSSR, blinded sample size reestimation; IPS, internal pilot study; PL, pseudolikelihood; SD, standard deviation

course of a previously defined time frame. A number of models have been proposed for such longitudinal count data. These range from integer times series as Markov chains^{3–5} and random effects or latent process approaches^{6–8} to other parametric^{9–11} and semi-parametric models.¹² In this paper we will adopt an approach based on a gamma frailty model introduced by Henderson and Shimakura⁶ and extended in Fiocco et al.⁸ The model allows for longitudinally dependent count data while maintaining a negative binomial marginal distribution, which is important for coping with the overdispersion that is often present. One favourable property of the gamma frailty model, which we will make use of, is the possibility of defining temporal trends while maintaining the desired marginal negative binomial distributions. The form of temporal trend is kept general in our approach so as to allow for a broad application of the techniques. Within the gamma frailty model, we are also able to process incomplete data sequences. This is an important property, as many sequences are still incomplete at interim analysis. Including such observations can substantially improve the precision of parameter estimates.¹³

Planning clinical trials includes sample size estimation. Sample size estimates are subject to unknown nuisance parameters, which have to be assumed from prior studies. The uncertainty introduced by nuisance parameters can be reduced by using blinded sample size reestimation (BSSR) techniques. These methods are applied after trial onset to estimate nuisance parameters on blinded data accrued to that point, without compromising trial integrity. Experimenters can then adapt the required sample size accordingly. Designs with internal pilot studies¹⁴ (IPS) for blinded sample size reestimation have been considered for several different statistical models and endpoints. For an overview we refer to Proschan^{15,16} and Friede and Kieser¹⁷ with more recent developments, considering count data, given in Schneider et al.,¹³ Friede and Schmidli^{18,19} and Asendorf et al.²⁰ Friede and Schmidli^{18,19} developed blinded sample size reestimation methods for univariate count data in two-arm placebo controlled trials, which are applied in experiments involving the number of relapses in MS. These methods were extended by Schneider et al.¹³ to consider incomplete observation sequences and time trends in event rates, in the context of recurrent event data. Asendorf et al.²⁰ extended the procedures for longitudinal count data, measured over regular time intervals, also considering incomplete observations but without trends in time. In this paper, we extend the methods of Asendorf et al.²⁰ by developing sample size estimation and reestimation techniques for longitudinal count data which can take into account temporal trends.

This paper is organised as follows. In Section 2 we provide context on MS trials and temporal trends. Section 3 introduces the statistical model upon which our estimation and testing procedures are based, the latter being explained in Section 4. In Section 5 we derive sample size estimation and reestimation procedures, and demonstrate these on MRI data from a recent study in MS. We conduct simulation studies to demonstrate the validity of the derived procedures in Section 6 and in Section 7 explore possibilities for adapting the procedures for small sample sizes. Section 8 provides discussion of the advantages and disadvantages of the proposed methods.

2 | CHALLENGES IN LESION COUNTS OF EARLY PHASE MS TRIALS

There is increasing evidence of a need for methods for the design and analysis of MS clinical trials in the presence of long or short term temporal trends. For example, a decreasing trend for annualized relapse rates has been shown across trials,^{21,22} which Stellmann et al.²³ consider responsible for an increase in the sample size of MS phase III trials, which they observed over the past 20 years. Nicholas et al.²² observe a decrease of the annualized relapse rates over follow-up, within trials. Given the close relationship between the number of relapses and lesion counts,²⁴ Stellmann et al.²³ suspect a similar trend for the number of new lesions. Another example is given in van Noort et al.,²⁵ who test for significantly decreasing lesion counts in a 36 week trial and Kappos et al.,²⁶ who show a trend in the proportion of patients free of lesions in a 6 month trial. Other examples in which temporal trend approaches will be useful include cumulative T1 lesions, as displayed for example in Comi et al.,²⁷ and trials with baseline to treatment designs.^{28,29}

Whenever planning sample sizes, some parameters have to be estimated from previous studies. Finding suitable reference studies for sample size planning can be challenging, especially in phase II MS trials, as the studies tend to be heterogeneous in their lesion types and evaluation methods. Further, as Chalavi et al.³⁰ summarise, the quality and evaluation of MRIs can depend on a number of factors including the scanner manufacturer, field strength, MRI protocol, scanner drift over time and employed analysis tools, all of which subsequently influence the number of lesion counts when MRI are examined by different laboratories. These factors lead to an increased uncertainty in sample size determination, when assuming parameters observed in previous studies apply also to the planned trial.

3 | STATISTICAL MODEL

The negative binomial distribution is commonly applied to overdispersed count data. It can be derived as a gamma frailty mixture of Poisson random variables. Serial correlation within a longitudinal count sequence can be modelled via a multivariate gamma distribution described by Henderson and Shimakura⁶ and extended for more general purposes in Fiocco et al.⁸ [To better understand the sample size reestimation procedures, we introduce the required notation, briefly explain the model and describe how general time trends can be incorporated. For more details on the estimation procedure and the gamma frailty model used here, we refer to Fiocco et al.⁸](#)

3.1 | The longitudinal gamma frailty model

The setting we consider is a two-arm randomised controlled study in which patients are observed at fixed time points over a specific time frame. To distinguish between the experimental group and the control group, we introduce the index $i = E, C$,

where E stands for the experimental group and C for the control group. Further, let $j = 1, \dots, n_i$ for $i = E, C$ denote the patient and $t = 1, \dots, T$ denote the time point at which observations are recorded. We begin by assuming that each patient has an associated unobservable multivariate gamma random variable

$$\mathbf{Z}_{ij} = (Z_{ij}^{(1)}, \dots, Z_{ij}^{(T)})^\top.$$

Each $Z_{ij}^{(t)}$ has marginal gamma distribution with mean one and variance η , which we denote as $Z_{ij}^{(t)} \sim \gamma(1, \eta^{-1})$. The within-patient frailties are not independent however, and $\text{Cor}(Z_{ij}^{(t)}, Z_{ij}^{(s)}) = \rho^{|s-t|}$. For information on how to generate random variables from a multivariate gamma distribution with this structure, we refer to Fiocco et al.⁸

The observed counts $X_{ij}^{(t)}$ are conditionally independent Poisson variables, given the frailties, which act multiplicatively upon the means. Thus,

$$X_{ij}^{(t)} | Z_{ij}^{(t)} \sim \text{Poisson}(\mu_i^{(t)} Z_{ij}^{(t)})$$

for some fixed group- and time-specific mean $\mu_i^{(t)}$. The marginal distribution is then negative binomial, $X_{ij}^{(t)} \sim NB(\mu_i^{(t)}, \eta)$ as required. Further, the within-patient counts are dependent, and following Fiocco et al.⁸ it can be shown that

$$\text{Cor}(X_{ij}^{(t)}, X_{ij}^{(s)}) = \rho^{|t-s|} \cdot \frac{\mu_i^{(t)} \mu_i^{(s)}}{\eta} \left/ \left(\sqrt{\frac{\mu_i^{(t)2}}{\eta} + \mu_i^{(t)}} \sqrt{\frac{\mu_i^{(s)2}}{\eta} + \mu_i^{(s)}} \right) \right.$$

The correlation structure of the observed counts has an autoregressive character inherited from the autoregressive frailty correlations, but the exact values depend upon the time trends as well as the time separation $|t - s|$.

3.2 | Incorporating trends

In principle each conditional mean $\mu_i^{(t)}$ can be estimated without further assumptions [and the proposed method is therefore quite flexible](#). However, in practice it is usually convenient to assume a lower-dimensional parametric form. We assume therefore that

$$\mu_i^{(t)} = f_i^{(t)}(\lambda) \quad i = E, C$$

for group-specific functions which are known up to a vector of *trend parameters* λ . For technical reasons we will encounter later, each $f_i^{(t)}(\cdot)$ is assumed to be a twice differentiable function with respect to the trend parameters.

For example, the standard time-constant model might be parametrised as $\lambda = (\lambda_1, \lambda_2)$ and

$$f_E^{(t)}(\lambda) = \exp(\lambda_1 + \lambda_2) \text{ and } f_C^{(t)}(\lambda) = \exp(\lambda_1).$$

This assumes that the effect of treatment is both constant and immediate. A natural extension is to $\lambda = (\lambda_1, \lambda_2, \lambda_3)$ and

$$f_E^{(t)}(\lambda) = \exp(\lambda_1 + t \cdot (\lambda_2 + \lambda_3)) \text{ and } f_C^{(t)}(\lambda) = \exp(\lambda_1 + t \cdot \lambda_2).$$

Counts at treatment onset now have the same mean, with the treatment effect growing with time. Figure 1 provides illustrations for both examples.

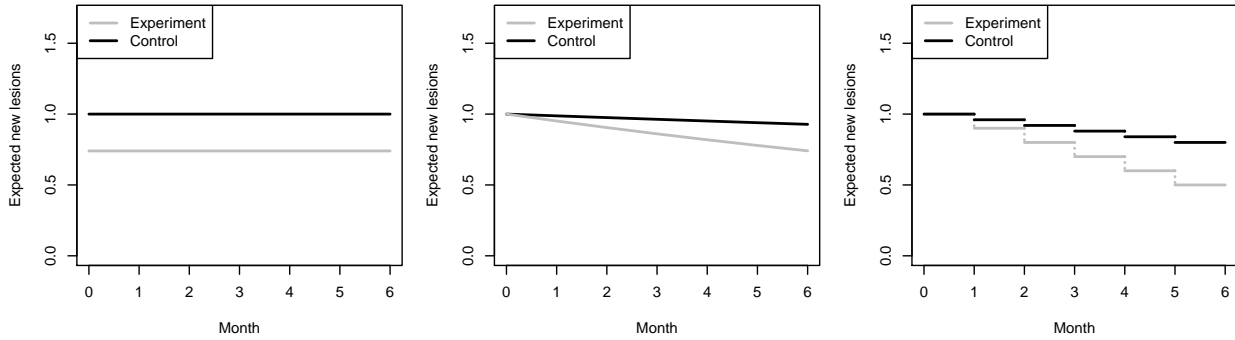


FIGURE 1 Illustrative examples of expected new lesion counts in monthly MRI data, given constant rates (left), an underlying exponential trend (center) and piecewise constant rates (right).

Note that, despite their appearance, the curves in the center panel of Figure 1 are not linear. Depending on the data situation, different trends might be more appealing. Because the functions $f_i^{(t)}(\cdot)$ need only be differentiable with respect to λ and not with respect to the time points t , there is a high flexibility when choosing other trends. This includes, for example, piecewise constant rates as displayed in the right panel of Figure 1, where arbitrary means at each time point can be defined, i.e. $f_i^{(t)}(\lambda) = \lambda_i^{(t)}$, with $\lambda = (\lambda_E^{(1)}, \dots, \lambda_C^{(T)})$ is restricted to the $2 \cdot T$ -dimensional space of positive real numbers. As trend specifications may differ depending on future applications, all results in the following sections are in terms of general functions $f_i^{(t)}(\lambda)$.

4 | FORMULATING AND TESTING HYPOTHESES

Prior to deriving sample size estimation and reestimation techniques, we give details on how the estimation and testing procedures proposed in Fiocco et al.⁸ are adapted to our setting.

4.1 | Parameter estimation and asymptotic properties

Given observations \mathbf{x} , we use composite likelihood methods for estimation and inference. The trend parameters λ and shape parameter η together determine the marginal distributions of counts. Hence we can use a pseudolikelihood (PL) approach with independence working assumption to obtain estimates $\hat{\lambda}$ and $\hat{\eta}$ by maximising

$$l(\eta, \lambda | \mathbf{x}) = \frac{1}{n_C + n_E} \sum_{i=E,C} \sum_{j=1}^{n_i} l_{ij}(\eta, \lambda | \mathbf{x}),$$

where

$$l_{ij}(\eta, \lambda | \mathbf{x}) = \sum_{t=1}^T \ln \left(P_{\text{NB}} \left(x_{ij}^{(t)}, f_i^{(t)}(\lambda), \eta \right) \right). \quad (1)$$

Here $P_{\text{NB}}(x, \mu, \eta)$ denotes the probability function of a negative binomial random variable with mean μ and variance $\mu^2/\eta + \mu$, see Appendix A.1 for details. An estimate $\hat{\rho}$ of the correlation parameter ρ can be attained at a second step by maximising the pairwise composite likelihood

$$cl(\rho | \hat{\eta}, \hat{\lambda}, \mathbf{x}) = \frac{1}{n_C + n_E} \sum_{i=E,C} \sum_{j=1}^{n_i} \sum_{t=1}^{T-1} \sum_{s=t+1}^T \ln(P_{\text{NB}}^{\text{pair}} \left(x_{ij}^{(t)}, x_{ij}^{(s)}; \rho, f_i^{(t)}(\hat{\lambda}), f_i^{(s)}(\hat{\lambda}) \right)), \quad (2)$$

where $P_{\text{NB}}^{\text{pair}}(\cdot, \cdot)$ denotes the bivariate probability function for counts at different times but on the same patient, see Appendix A.2 for details.

Attaining standard errors for the PL estimates proves to be slightly more difficult, as the observations are not independent over time points. Here, we focus on attaining standard errors for $\hat{\eta}$ and $\hat{\lambda}$, referring to Fiocco et al.⁸ for estimating standard errors for $\hat{\rho}$. Because we maximise the pseudolikelihood with an independence working correlation, the associated Fisher information does not lead to consistent estimation of standard errors of the PL estimates. Instead, we consider the sandwich estimator, which requires the first and second derivatives of the log-likelihood contributions, namely

$$\nabla_{(\eta, \lambda)} l(\eta, \lambda | \mathbf{x}) = \left(\frac{\partial l(\eta, \lambda | \mathbf{x})}{\partial \eta}, \frac{\partial l(\eta, \lambda | \mathbf{x})}{\partial \lambda_1}, \dots, \frac{\partial l(\eta, \lambda | \mathbf{x})}{\partial \lambda_D} \right)^\top$$

and

$$\nabla_{(\eta, \lambda)(\eta, \lambda)} l(\eta, \lambda | \mathbf{x}) = \begin{pmatrix} \frac{\partial^2 l(\eta, \lambda | \mathbf{x})}{\partial \eta^2} & \left(\frac{\partial^2 l(\eta, \lambda | \mathbf{x})}{\partial \eta \partial \lambda_d} \right)_{d=1, \dots, D}^\top \\ \left(\frac{\partial^2 l(\eta, \lambda | \mathbf{x})}{\partial \eta \partial \lambda_d} \right)_{d=1, \dots, D} & \left(\frac{\partial^2 l(\eta, \lambda | \mathbf{x})}{\partial \lambda_d \partial \lambda_e} \right)_{d=1, \dots, D, e=1, \dots, D} \end{pmatrix}.$$

Details on the calculation of these terms are provided in Appendix B and C. For deriving asymptotic properties of $\hat{\lambda}$ and $\hat{\eta}$ we require the expected values of the Hessian matrix and expected values of the outer gradient matrix, denoted by

$$H = \frac{1}{n_C + n_E} \sum_{i=E,C} \sum_{j=1}^{n_i} E \left[\nabla_{(\eta, \lambda)(\eta, \lambda)} l_{ij}(\eta, \lambda | \mathbf{x}) \right] \quad (3)$$

and

$$J = \frac{1}{n_C + n_E} \sum_{i=E,C} \sum_{j=1}^{n_i} E \left[\nabla_{(\eta, \lambda)} l_{ij}(\eta, \lambda | \mathbf{x}) \cdot \nabla_{(\eta, \lambda)} l_{ij}(\eta, \lambda | \mathbf{x})^\top \right] \quad (4)$$

respectively. Further information on calculating H and J is in Appendix D. Standard asymptotic theory on likelihood estimators yields

$$\sqrt{n_C + n_E} \cdot \left((\hat{\eta}, \hat{\lambda}) - (\eta, \lambda) \right) \rightarrow N(0, H^{-1} J H^{-1}) \quad (5)$$

for $n_C, n_E \rightarrow \infty$ with $k = n_E/n_C$ constant, and where convergence is in distribution. See Appendix E for details. With the help of this asymptotic property, we will consider suitable hypotheses and derive applicable test statistics.

4.2 | Considered hypotheses

Throughout the remainder of this paper we consider testing hypotheses of the form $H_0 : h(\eta, \lambda) \geq h_0$ where the function $h : \mathbb{R}^{D+1} \rightarrow \mathbb{R}$ is differentiable and monotone in all dimensions. Reconsidering the simple examples from Section 3.2, a hypothesis for the constant trend could be $H_0 : \lambda_2 = 0$, for the exponential trend it could be $H_0 : \lambda_3 = 0$, and for the piecewise constant model it might be $H_0 : \lambda_E^{(i)} = \kappa \lambda_C^{(i)}$ for some time-fixed constant κ . The multivariate delta method in combination with the asymptotic result in (5) yields

$$\sqrt{n_C + n_E} \cdot \left(h(\hat{\eta}, \hat{\lambda}) - h(\eta, \lambda) \right) \xrightarrow{n_E, n_C \rightarrow \infty} N \left(0, \nabla_{(\eta, \lambda)} h(\eta, \lambda)^\top H^{-1} J H^{-1} \nabla_{(\eta, \lambda)} h(\eta, \lambda) \right)$$

with $k = n_E/n_C$ constant. Using estimators

$$\hat{H} = \frac{1}{n_C + n_E} \sum_{i=E, C} \sum_{j=1}^{n_i} \nabla_{(\eta, \lambda)(\eta, \lambda)} l_{ij}(\hat{\eta}, \hat{\lambda} | \mathbf{x}) \quad (6)$$

and

$$\hat{J} = \frac{1}{n_C + n_E} \sum_{i=E, C} \sum_{j=1}^{n_i} \nabla_{(\eta, \lambda)} l_{ij}(\hat{\eta}, \hat{\lambda} | \mathbf{x}) \cdot \nabla_{(\eta, \lambda)} l_{ij}(\hat{\eta}, \hat{\lambda} | \mathbf{x})^\top, \quad (7)$$

we can derive a Wald-Type test statistic Z which asymptotically follows a standard normal distribution

$$Z = \sqrt{n_C + n_E} \cdot \frac{h(\hat{\eta}, \hat{\lambda}) - h_0}{\sqrt{\nabla_{(\eta, \lambda)} h(\hat{\eta}, \hat{\lambda})^\top (\hat{H}^{-1} \hat{J} \hat{H}^{-1}) \nabla_{(\eta, \lambda)} h(\hat{\eta}, \hat{\lambda})}} \overset{\cdot}{\sim} N(0, 1). \quad (8)$$

On the basis of the test statistic Z and its asymptotic properties, we can derive a sample size formula. For improved readability in upcoming sections, we denote the standard error of $h(\hat{\eta}, \hat{\lambda})$ by

$$\sigma = \sqrt{\nabla_{(\eta, \lambda)} h(\eta, \lambda)^\top (H^{-1} J H^{-1}) \nabla_{(\eta, \lambda)} h(\eta, \lambda)}$$

and its plug-in estimator by $\hat{\sigma}$.

5 | SAMPLE SIZE CALCULATION AND BLINDED REESTIMATION

We first derive a sample size formula for the fixed design and describe blinded sample size reestimation procedures. The application is then demonstrated on lesion counts from a recent study in MS.

5.1 | Sample size formula for fixed designs

Using the asymptotic distribution given in (8), we can derive an approximate formula for calculating the sample size required for rejecting the null hypothesis H_0 with probability $1 - \beta$ at significance level α , given an effect size $\theta^* = h(\eta^*, \lambda^*) - h_0$. The

sample size of the control group is given by

$$n_C = \frac{(z_\beta + z_\alpha)^2 \cdot \sigma^2}{(1 + k) \cdot \theta^{*2}} \quad (9)$$

where z_γ denotes the γ -quantiles of the standard normal distribution. The sample size of the experimental group is given by $n_E = k \cdot n_C$ with treatment group allocation $k : 1$ for experimental to control. From the sample size formula, we see that the choice of power, significance level, expected effect size and sample size allocation factor k influence the sample size. All of these are known and can be planned within a study. What is not known and cannot be chosen, is the nuisance parameter σ^2 , which in turn can be calculated using the expected trend parameters λ , shape parameter η , correlation parameter ρ and number of time points T , as described in Section 4.2. We use simulations to investigate the influence of these nuisance parameters on sample size in Section 6.2.

5.2 | Blinded sample size reestimation

Blinded sample size reestimation is a three step procedure. First, the study parameters, relevant effect size and nuisance parameters are defined and an initial sample size is calculated using a sample size formula. Then, after the trial has started and a proportion of participants has been recruited into the trial, the sample size is reestimated using available blinded data, i.e. all data available at the interim time point, without information on the treatment allocation and so without compromising the trial integrity.³¹ Based on the reestimated sample size, recruitment of participants is adjusted and the final analysis is performed on the reestimated number of participants. Summarizing, we perform the following steps:

1. Calculate an initial sample size estimate $n_C^{(0)}$ and $n_E^{(0)} = k \cdot n_C^{(0)}$ for the control and experimental group, respectively, based on design options and the smallest clinically meaningful effect ($k, T, \theta^*, \alpha, \beta$) and assumptions on nuisance parameters (λ, η, ρ).
2. At a suitable interim time, perform blinded sample size reestimation to obtain sample size reestimates $n_C^{(1)}$ and $n_E^{(1)}$ based on estimates of nuisance parameters from blinded data of patients already recruited into the control and experimental group.
3. Perform the final analysis based on $\max(m_C^{(0)}, n_C^{(1)})$ participants for the control group and $\max(m_E^{(0)}, n_E^{(1)})$ for the experimental group, where $m_i^{(0)}$ denotes the number of patients enrolled by interim.

One popular method of blinded estimation for count data is given through the lumping approach,^{18,19,32} in which all blinded observations are treated as if they were from the same negative binomial distribution. However, we will make use of a different possibility, applied in Asendorf et al.,²⁰ which is given by treating blinded observations as observations from a mixture of two negative binomial distributions. The two approaches have been shown to give similar results.²⁰ Thus we estimate the parameters λ and η in a blinded fashion by maximising the following pseudolikelihood from a mixture of negative binomial distributions:

$$l_{\text{blind}}(\eta, \lambda | \mathbf{x}) = \frac{1}{n_C + n_E} \sum_{j=1}^{n_E + n_C} \sum_{t=1}^T \ln \left(\frac{k}{1+k} \cdot \text{P}_{\text{NB}} \left(x_j^{(t)}, f_E^{(t)}(\lambda), \eta \right) + \frac{1}{1+k} \cdot \text{P}_{\text{NB}} \left(x_j^{(t)}, f_C^{(t)}(\lambda), \eta \right) \right). \quad (10)$$

Whether blinded estimation of nuisance parameters is possible through a mixture distribution depends on how the group specific trend parameters are interrelated. Usually, the group specific trend parameters are interrelated by assuming the clinically relevant effect θ^* to be true. For illustrative purposes, consider the example of constant trends. Assuming that the clinically relevant effect holds implies that $\lambda_2 = \theta^*$. Therefore, there exists an invertible function $g(\cdot)$ such that $g(\lambda_C) = g(\lambda_1) = g(\lambda_1, \theta^*) = \lambda_E$, where λ_i denotes a vector of group specific trend parameters. More generally, for the blinded estimation of nuisance parameters to be reliable, such a function $g(\cdot)$ needs to exist. If it does not exist naturally, assumptions on the study design may need to be imposed to guarantee its existence.

Estimating the correlation parameter ρ proves to be slightly more difficult: we use a two stage procedure similar to that presented Section 4.1. In the first stage, we calculate PL estimates for $\hat{\lambda}$ and $\hat{\eta}$ as presented above. In the second stage, we maximise with respect to ρ a composite mixture of bivariate probability functions. This is

$$cI_{\text{blind}}(\rho|\hat{\eta}, \hat{\lambda}, \mathbf{x}) = \frac{1}{n_C + n_E} \sum_{j=1}^{n_E+n_C} \sum_{t=1}^{T-1} \sum_{s=t+1}^T \ln \left(\frac{k}{1+k} \cdot P_{\text{NB}}^{\text{pair}} \left(x_j^{(t)}, x_j^{(s)}; \rho, f_E^{(t)}(\hat{\lambda}), f_E^{(s)}(\hat{\lambda}), \hat{\eta} \right) \right. \\ \left. + \frac{1}{1+k} \cdot P_{\text{NB}}^{\text{pair}} \left(x_j^{(t)}, x_j^{(s)}; \rho, f_C^{(t)}(\hat{\lambda}), f_C^{(s)}(\hat{\lambda}), \hat{\eta} \right) \right). \quad (11)$$

The estimates $\hat{\lambda}$, $\hat{\eta}$ and $\hat{\rho}$ attained from the blinded data are then used to calculate $\hat{\sigma}^2$, which is plugged into the sample size formula to attain a reestimated sample size.

5.3 | Incorporating incomplete observations

A frequently encountered problem when performing blinded reestimation on longitudinal data, is the case of incomplete follow up. In most clinical trials, a certain fraction of patients will not have been fully observed when the blinded sample size reestimation is conducted, resulting in administrative missingness at the reestimation time. [This type of missingness can be considered to be completely missing at random.](#) Inclusion of partial data from these patients however, can lead to an information gain and therefore more precise reestimation of nuisance parameters and the required sample size. Incorporating observations with incomplete follow up into the blinded sample size reestimation is possible by replacing the fixed follow up time T with patient-specific follow up times T_j in both likelihoods for blinded estimation, (10) and (11).

The problem of incomplete observations may also persist in the final analysis. Under the assumption that observations are missing completely at random, the same technique described for the blinded sample size reestimation procedure can be applied for the final analysis. Instead of taking a fixed follow up time T in the pseudolikelihood (1) and composite likelihood (2), each patient is given a specific follow up time T_j . Subsequent changes to the calculations of the sandwich estimator are straightforward.

5.4 | Application of the BSSR Procedure

For illustration we consider data from Fernandez et al.,³³ available online as supplementary material to their publication. In this MS trial, active T1 lesions were measured at three time points (at baseline, 6 months after randomization, 12 months after randomization) on patients who were administered different doses of autologous Adipose-derived mesenchymal stem cells (AdMSC). We consider the placebo group ($n = 11$) and the low-dose group ($n = 10$), in the latter 1×10^6 cells/kg autologous AdMSC were administered to the patients. Figure 2 illustrates the individual patient data of these two groups (center and right panel), as well as expected lesion counts from the observed data and the corresponding fit with the exponential trend model (left panel).

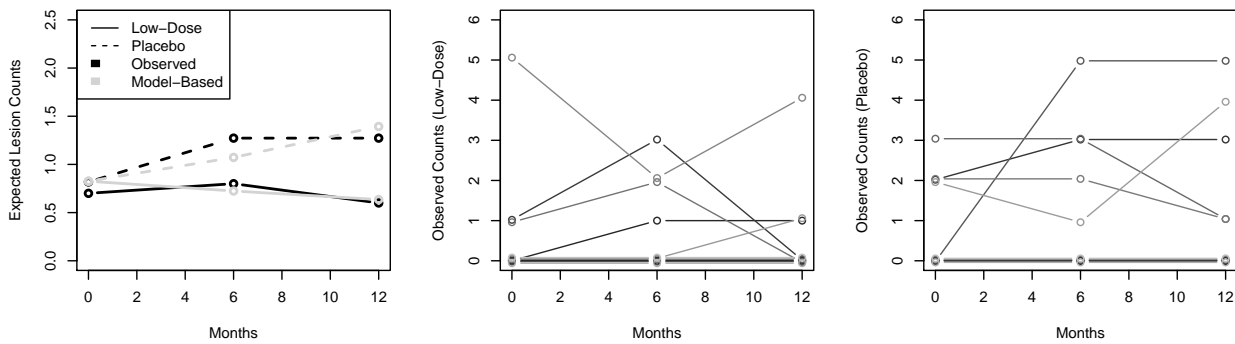


FIGURE 2 Expected lesion counts of the observed data, estimated by negative binomial regression and the corresponding fit with the exponential trend model (left). Observed individual lesion counts of the low-dose treatment group (center) and of the placebo group (right).

Figure 2 suggests that lesion counts are slightly decreasing with time in the low-dose group, and increasing in the placebo group. The shape parameters attained by negative binomial regression on the observed data and fitting the exponential model are estimated as 0.4955 and 0.4954, respectively.

For the purpose of illustration, we will pool the data across the treatment groups and demonstrate how a sample size review could be conducted assuming that this is interim data of an ongoing trial. We will assume an exponential trend with treatment effect $\theta^* = \lambda_3 = -0.70$ and plan to reject the null hypothesis $H_0 : \lambda_3 \geq 0$ with a power of at least 80% at nominal level $\alpha = 2.5\%$. Furthermore, the sample size allocation is $k = 1$ and the number of time points is $T = 3$. Then, the blinded sample size reestimation procedure yields a required sample size of 31 patients per group. To meet these power requirements, the trial would therefore have to be extended from 21 current patients to a total of 62 patients. For more details on the performed blinded sample size reestimation including R code, we refer to Appendix F.

6 | OPERATING CHARACTERISTICS OF THE BLINDED SAMPLE SIZE REESTIMATION PROCEDURE

In this section we investigate the characteristics of the presented BSSR procedure by comparing it to a fixed design without sample size reestimation. We follow the three step procedure described in Section 5 and compare the results to having only performed an initial sample size estimation, i.e. the first step of this procedure. We compare type I error and power of the procedures and demonstrate the benefit of incorporating incomplete observations. Table 1 gives an overview of the simulation settings considered in this section. We have concentrated on values that are realistic for MRI phase II studies in relapsing- and remitting MS.

TABLE 1 Overview of parameters considered in simulation settings for superiority tests. The IPS is conducted when complete follow-up is available on $n_C^{(0)}/2$ and $n_E^{(0)}/2$ patients. For each trend type, simulations are conducted with all possible combinations of parameters. Note, that for the constant and exponential trend the relevant effect θ^* corresponds to λ_2 and λ_3 , respectively.

Trend	Parameter		Values
Constant	Intercept rate	λ_1	-1.0, -0.5, 0.0, 0.5, 1.0
	Shape parameter	η	0.5, 1.0, 1.5
	Correlation parameter	ρ	0.0, 0.5, 0.9
	Null hypothesis	h_0	0
	Sample size Allocation	k	1
	Number of time points	T	6
	Nominal level (one-sided)	α	0.025
	Power	$1 - \beta$	0.8
	Relevant effect (log-scale)	θ^*	-0.3
Exponential	Intercept rate	λ_1	-1.0, -0.5, 0.0, 0.5, 1.0
	Slope rate	λ_2	-0.083, 0.0, 0.083
	Shape parameter	η	0.5, 1.0, 1.5
	Correlation parameter	ρ	0.0, 0.5, 0.9
	Null hypothesis	h_0	0
	Sample size Allocation	k	1
	Number of time points	T	7
	Nominal level (one-sided)	α	0.025
	Power	$1 - \beta$	0.8
	Relevant effect (log-scale)	θ^*	-0.05

6.1 | Type I error

In a first simulation study, we investigate if adjusting the sample size, through blinded sample size reestimation, has an impact on the type I error rate, compared to the fixed design. We perform all simulations separately for constant rates and exponential trend.

6.1.1 | Constant rates

In the presence of constant rates, we are mainly interested in testing the hypothesis $H_0 : \lambda_2 \geq 0$ vs. $H_A : \lambda_2 < 0$. The nuisance parameters were expected to be $\lambda_1 = 0$, $\eta = 1$ and $\rho = 0.5$, effectively corresponding to one new monthly lesion per patient in the control group and 0.74 new monthly lesions per patient in the experimental group. With these assumptions, the initial sample size estimate for the control group is $n_C^{(0)} = 102$.

Additionally to these simulation runs, we also investigate the type I error rates when testing for non-inferiority. Situations exist in which the type I error rates are inflated depending on the non-inferiority margin.³⁴ To investigate, we also perform simulations in which we consider the hypothesis $H_0 : \lambda_2 \geq \delta$ vs. $H_A : \lambda_2 < \delta$ for some non-inferiority margin $\delta > 0$, while assuming that experiment and control group have the same underlying effect. Thus, for the non-inferiority trials, we assume $\lambda_2^* = 0$ and set $\delta = h_0 = 0.2$, resulting in $\theta^* = -0.2$, while all other parameters remain as displayed in Table 1 .

Having calculated the sample size for the fixed design, we simulate data under the null hypothesis, but with different than expected nuisance parameters. Blinded sample size reestimates are attained from half of the initially simulated fixed data set and the data set is extended as described in Section 5. The settings we evaluated are given through all possible combinations of parameter values displayed in Table 1 , resulting in a total of 90 different simulation settings, with actually required sample sizes ranging from 30 to 357 patients. Rejection probabilities for both the fixed design and blinded sample size reestimation, as well as non-inferiority tests, are displayed in Figure 3 .

The type I error rate of the fixed design seems to be moderately inflated, a property which was observed in similar settings elsewhere.^{35,36} A mixed effect multiple logistic regression was performed to assess any correlation between parameters and type I error rates. Fixed effects were λ_1 , η , ρ , the type of hypothesis (non-inferiority vs. superiority) and the procedure (fixed vs. BSSR). The data set, on which the procedure is performed, is modeled as a random effect, as it is partially shared in both procedures. Only the type of hypothesis was found to have a significant influence on the type I error ($p < 0.0001$), other fixed effects were not statistically significant. The BSSR procedure does not seem to show higher type I error rates than the fixed design (Figure 3). When testing for non-inferiority, the BSSR procedure did not inflate the type I error rate. Type I error rates are not as high as observed with superiority tests, mainly due to the lower effect size and therefore higher sample sizes in the non-inferiority simulation runs.

6.1.2 | Exponential trend

When there is an underlying exponential trend in the data, the main interest lies in testing the hypothesis $H_0 : \lambda_3 \geq 0$ vs. $H_A : \lambda_3 < 0$. Following the notation from Section 4.2, we therefore have $h(\eta, \lambda) = \lambda_3$ and $h_0 = 0$. The nuisance parameters were expected to be $\lambda_1 = 0$, $\lambda_2 = 0$, $\eta = 1$ and $\rho = 0.5$. With these assumptions, the sample size for the control group is $n_C^{(0)} = 229$.

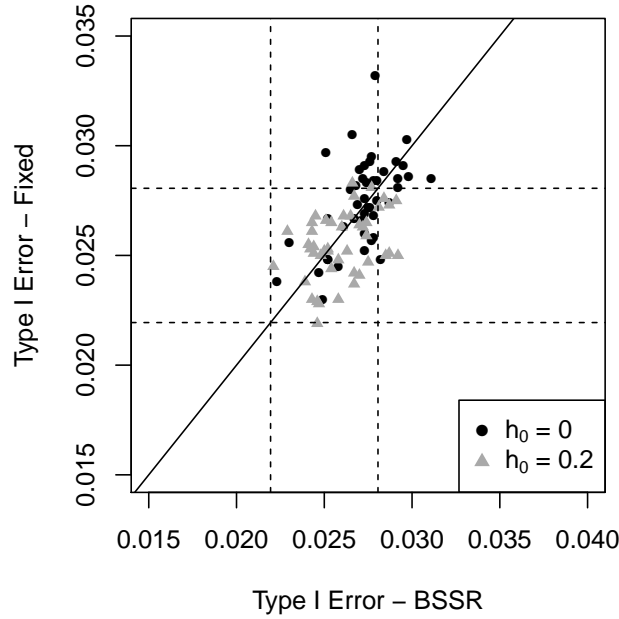


FIGURE 3 Comparison of type I error rates under constant trend and two different null hypotheses, i.e. superiority ($h_0 = 0$) and non-inferiority ($h_0 = 0.2$). The nominal level is set at $\alpha = 0.025$ (one-sided). Nominal 95% random intervals are given for orientation. Each dot corresponds to one of the settings from Table 1 . Results from non-inferiority tests are coloured grey.

We again simulate the data with nuisance parameter values different from the expected values and no treatment effect. The settings we evaluate are given through all possible combinations of parameter values as displayed in Table 1 , resulting in a total of 135 distinct settings with true required sample sizes ranging from 63 to 875 patients. Type I error rates for both the fixed design and blinded sample size reestimation are displayed in Figure 4 .

The type I error rate of the fixed design again seems to be slightly above nominal level. The type I error rate of the BSSR procedure is slightly lower, but very similar to that of the fixed design. A mixed effect multiple logistic regression between type I error rates and parameter setting, with λ_1 , λ_2 , η , ρ and procedure (fixed vs. BSSR) as fixed effects and the data set as random effect, did not reveal any correlation between type I error rates and specific parameter settings. Overall, the BSSR procedure, which uses blinded data from a running trial to reestimate nuisance parameters, does not inflate the type I error rate beyond levels seen in fixed designs and can therefore be used without any adjustment of the nominal level.

6.2 | Power and sample size

The main purpose of considering a blinded sample size reestimation lies in correcting the sample size to achieve a predefined power, when nuisance parameters are misspecified and the clinical effect assumed is present. Following Birkett and Day,³⁷ we allow the reestimated sample size to be lower than the initially planned sample size. In practice, however, this is sometimes

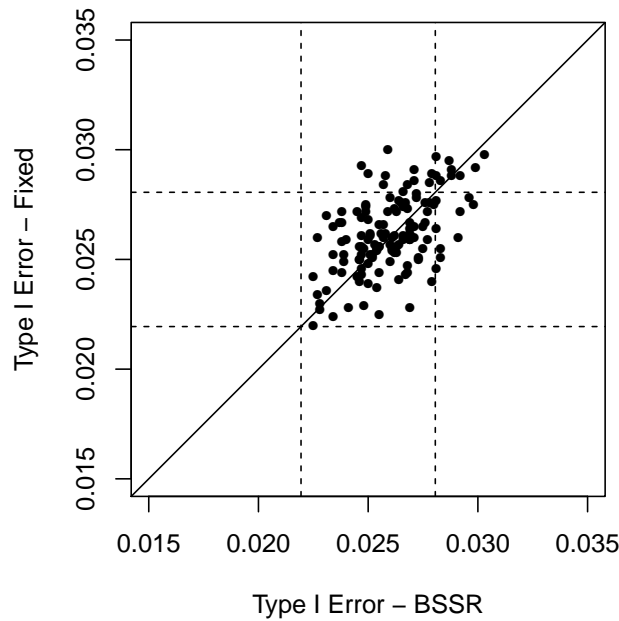


FIGURE 4 Comparison of type I error rates under exponential trend. The nominal level is set at $\alpha = 0.025$ (one-sided). Nominal 95% random intervals are given for orientation. Each dot corresponds to one of the settings from Table 1 .

discouraged. Estimates of nuisance parameters are always subject to variance, especially early in the study. Therefore, allowing for a decrease in the sample size increases the risk of an underpowered study, potentially caused merely by the variability of nuisance parameters estimates.³¹

6.2.1 | Constant rates

Similarly to the type I error simulation, we perform an initial sample size estimation by choosing study parameters and relevant effect size as in Table 1 . The nuisance parameters are again expected to be $\lambda_1 = 0$, $\eta = 1$ and $\rho = 0.5$, leading to $n_C^{(0)} = 102$.

The data are now simulated with the relevant effect size being present, i.e. $\lambda_2 = -0.3$. However, the nuisance parameters are varied away from the a priori expected values in a sequential manner. First we vary λ_1 and simulate data with $\lambda_1 = -1, \dots, 1$ in steps of 0.1 while keeping ρ and η equal to the expected values. Next we vary $\rho = 0, \dots, 1$ in steps of 0.1 while keeping λ_1 and η constant. Finally, $\eta = 0.5, \dots, 1.5$ is varied in steps of 0.1 while the other two nuisance parameters are kept fixed. In each setting, a sample size reestimation is performed aimed at correcting the sample size to achieve a power of 80% at the end of the study. The results are displayed in Figure 5 .

For all of the nuisance parameters, a misspecification leads to the fixed design not achieving the desired power. The sample size estimated within the BSSR procedure on the other hand corrects the sample size appropriately to attain the desired power.

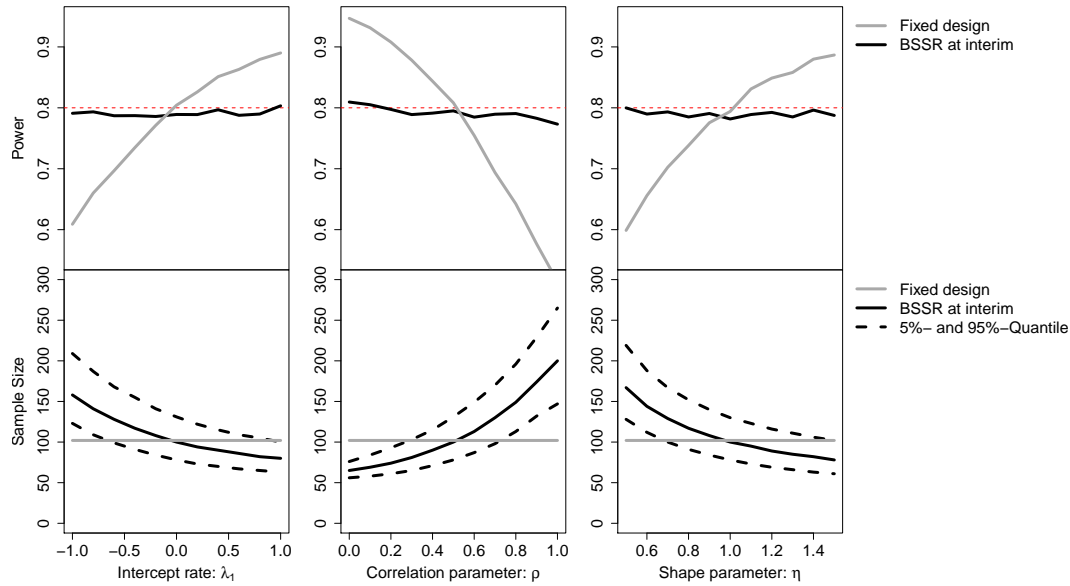


FIGURE 5 Power and sample size of the fixed design and the BSSR procedure when nuisance parameters are misspecified under constant rates.

6.2.2 | Exponential trend

For the exponential trend we also compare the fixed design to the BSSR procedure in terms of power and sample size. The study parameters and relevant effect size are chosen as displayed in Table 1. The nuisance parameters were expected to be $\lambda_1 = 0$, $\lambda_2 = 0$, $\eta = 1$ and $\rho = 0.5$, which gave $n_C^{(0)} = 229$. Again the data are simulated with the treatment effect present but nuisance parameters allowed to vary sequentially, as for the previous section. The power comparisons between the fixed design and BSSR procedure are presented in Figure 6.

For all nuisance parameters, the fixed design does not meet the power requirements when there is misspecification. The BSSR procedure corrects the sample size when nuisance parameters have been misspecified and achieves the desired power of 80% in the final analysis.

6.3 | Incomplete observations

To investigate the effect of incomplete observations being present at interim analysis, we consider the setting described in Section 6.2.1 with an initial sample size estimate of $n_C^{(0)} = 102$ in each group. Let us assume that each month a total of 20 patients, 10 in each group, are enrolled to the trial. The follow-up time per patient is six months (monthly MRI scans), therefore the trial will be about 16 months long with an enrollment phase of 10 months. We would like to conduct a sample size reestimation before enrollment ends, for example six months after trial onset. At that time we have only 10 observations in each group which are

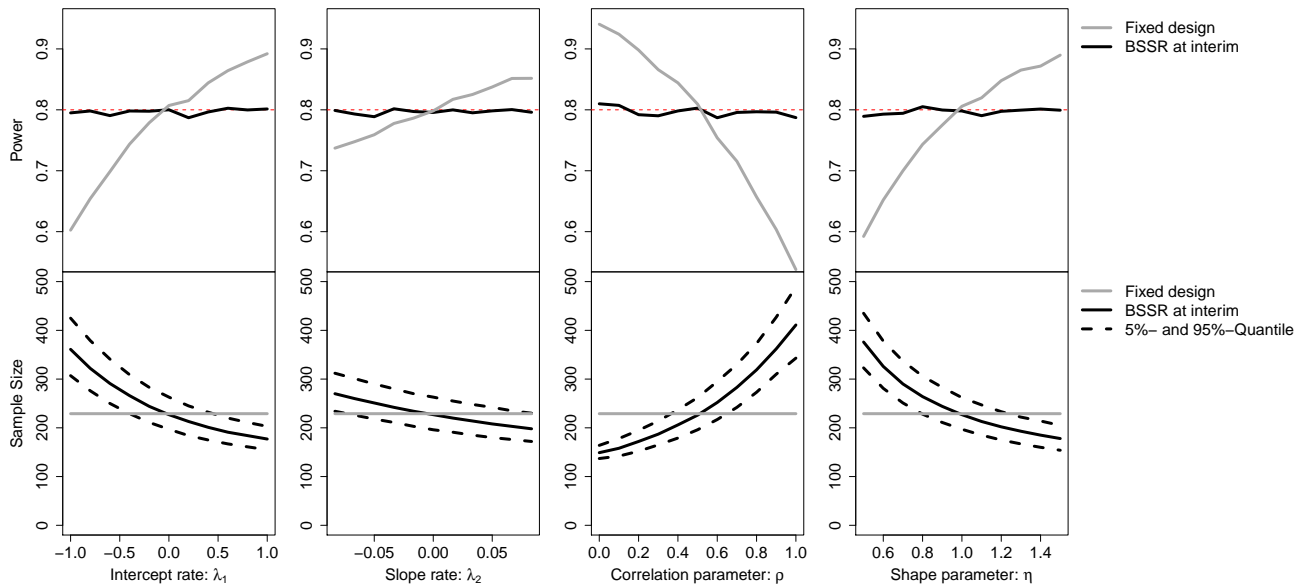


FIGURE 6 Power and sample size of the fixed design and the BSSR procedure when nuisance parameters are misspecified under an exponential trend.

complete, although 60 patients per group have been enrolled and can provide data. Table 2 shows the benefit of incorporating the incomplete observations in comparison to only considering complete sequences.

TABLE 2 Mean and standard deviation (SD) of sample size estimates at different review time points, including incomplete observations and using only complete observations.

Month of review	Including incomplete observations		Complete observations only	
	Mean	SD	Mean	SD
6	102.3	20.3	100.5	34.8
7	102.2	17.6	101.1	25.3
8	101.8	15.9	101.2	20.7
9	102.2	14.7	101.8	18.2
10	102.0	13.5	101.8	16.3

We see that the mean sample size estimate is not affected by including incomplete observations but the precision is markedly improved. This effect is strongest in early sample size reviews, when the number of incomplete observations is high in comparison to the number of complete observations. For instance, by including incomplete observations, a standard deviation of about 20 can already be achieved at month 6, as opposed to month 8 with complete observations only.

7 | AN EXTENSION FOR SMALL SAMPLE SIZES

The simulations in Sections 6.1.1 and 6.1.2 show the fixed design to have type I error rates slightly above the nominal level. The observed inflation was rooted in the testing procedure and was not specific to the blinded sample size reestimation procedure. While the inflation is partially explained through a relatively low sample size implied by the simulation settings, methods have been studied which might lead to improved control of the type I error rate.^{38,39} To improve type I error rate control we adapt an approach as in Mütze et al.,³⁹ in which the variance of the effect size is estimated under the null hypothesis by means of restricted maximum likelihood estimation. This concept, of restricting the variance estimator to the null hypothesis for improved inference results, has been studied previously, including Mielke and Munk⁴⁰ and references given therein. Transferred to the setting considered here, this implies calculating the expected Hessian and outer gradient matrix analogously to (3) and (4), but restricted to the one-sided null hypothesis as stated in Section 4.2. The restricted expected Hessian and restricted outer gradient matrix are denoted as $H^{(r)}$ and $J^{(r)}$, respectively. The resulting sandwich estimator is referred to as the restricted variance estimate.

Calculation of the restricted variance estimate is performed in three steps. First, parameter estimates $\hat{\eta}_r$ and $\hat{\lambda}_r$ for the shape parameter η and trend parameters λ , respectively, are attained by restricted pseudolikelihood estimation. Note that the optimization is conducted over the half-open space $\{(\eta, \lambda) \in \mathbb{R}^{D+1} | h(\eta, \lambda) \geq h_0\}$ and not only over the rim of the hypothesis. Secondly, these restricted estimates are plugged into the composite likelihood to attain a restricted estimate $\hat{\rho}_r$ of the correlation parameter ρ . Thirdly, the restricted estimates, $\hat{\eta}_r$, $\hat{\lambda}_r$ and $\hat{\rho}_r$, are plugged into equations (3) and (4) to attain the restricted expected Hessian $H^{(r)}$ and restricted outer gradient matrix $J^{(r)}$.

To illustrate the improvements of the proposed extension for small sample sizes, we reconsider the simulation of superiority trials from Section 6.1.1. We take the same parameter values except we consider two different effect sizes, $\theta^* = -0.3$, as previously considered, and also $\theta^* = -0.6$, which reduces the original sample size from 102 to 27 observations per group. Figure 7 compares results in the fixed design, with inference is based on the original and restricted variance estimator.

At the smaller sample size there is a clear inflation of the type I error rate when the unrestricted method is used, which is corrected by the restricted variance procedure.

8 | DISCUSSION

Motivated by MRI lesion counts, we considered a gamma frailty model for longitudinal count data with overdispersion, dependence and possible time-trends in mean values.⁸ The model maintains a marginal negative binomial distribution. Correlations between observations are implied by correlations between the multivariate gamma frailty components.

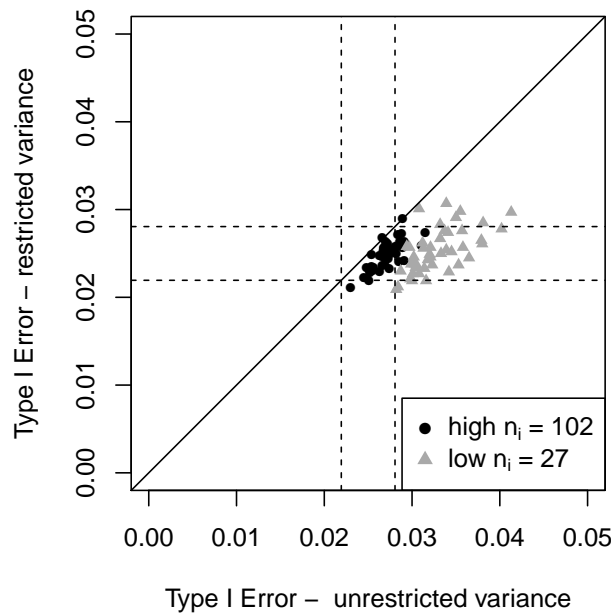


FIGURE 7 Comparison of type I error rates when calculating the restricted variance estimate compared to inference based on the unrestricted variance estimate, as described in Section 4.2. Each dot corresponds to one of the settings for constant rates from Table 1. The nominal level for all simulation settings is $\alpha = 2.5\%$ (one-sided). Black dots and gray triangles correspond to a sample size of 102 ($\theta^* = -0.3$) and 27 ($\theta^* = -0.6$) observations per group, respectively.

Within the gamma frailty model, using pseudolikelihood estimators and their asymptotic properties, a Wald-Type test statistic was derived for testing one-sided and two-sided hypotheses of trend parameters. The modelling approach allows for more precise hypothesis formulation and inference than more simplistic approaches, commonly used in MS studies. The Wald-Type test statistic was used to derive a sample size formula. Although the sample size can be calculated numerically, [a more detailed expression giving insights on the influence of nuisance parameters on the standard error \$\sigma\$ could not be found. The difficulty in attaining a more detailed expression lies in formulating a general inverse of the expected Fisher information matrix for the sandwich estimator, as well as analytic calculations of specific terms of the expected Fisher information, e.g. \$E\[\psi_1\(x_{ij}^{\(t\)} + \eta\)\]\$, for which only numerical approximations exist.](#) Nevertheless, the influence of nuisance parameters on the sample size was illustrated in subsequent simulation results.

On the basis of the sample size formula, techniques for blinded sample size reestimation were developed. [Here we adopted the so-called mixture approach, opposed to the simpler lumping approach, which may also have been possible. The latter could potentially be implemented using generalized estimating equations, as correspondingly done in the context of longitudinal continuous data.](#)⁴¹ The techniques developed here were demonstrated for two different trends and shown to correct the sample size to give a power close to the target when nuisance parameters were a priori misspecified. Trend specifications were kept general

throughout the paper, so that other trends can easily be adopted into the presented framework, facilitating a broader application of the presented methods. The possibility of sample size estimation and reestimation for longitudinal negative binomial data with time trends can increase the efficiency of clinical trials with such endpoints. Authorities accept blinded sample size reestimation as a familiar design method which does not introduce statistical bias while maintaining Type I error control, and should therefore generally be considered for most studies.³¹

When planning a study, the optimal timing of a sample size review is an important aspect to consider. If the review is carried out too early, the variability of nuisance parameter estimates is high and with it the variability of the resulting sample size. High variability at early time points is also apparent in Table 2. Low variability of the reestimated sample size is always advantageous and therefore, an interim analysis at a late time point is tempting. However, if the sample size review is conducted very late, then the sample size may already be larger than required. Considering that the timing is a trade off between the variability of the reestimated sample size and risking a too high sample size, a number of approaches have been proposed for timing the sample size review. A summary of these, in the context of sample size reviews with recurrent event outcomes, can be found in Section 6 of Friede and Schmidli.¹⁸

We have considered the incorporation of incomplete responses arising for administrative purposes and realistically considered to be missing completely at random. Our method can also be adapted to situations where data are missing at random (MAR), in which case we would replace our two-stage estimation approach by simultaneous estimation of all parameters from the pairwise composite log-likelihoods (2) and (11) as appropriate. Maximum likelihood estimation is consistent under MAR and this property is inherited by the pairwise version provided that all parameters can be identified from the bivariate distributions, which is indeed the case for our model. Simultaneous estimation of all parameters may, however, substantially increase computation times.

The derived sample size reestimation procedures are based on the longitudinal gamma frailty model from Fiocco et al.⁸ Consequently, the procedures are limited to this model and may be inaccurate otherwise. Because the longitudinal gamma frailty model, which the sample size reestimation procedure depends on, cannot be fitted using standard software, the presented inference, sample size estimation and blinded sample size reestimation methods have been coded in R and made available on CRAN within the package `spass`.⁴² A detailed example using data from Fernandez et al.³³ is given in the Appendix F. Both variance estimators considered are implemented, namely using the observed sandwich estimator as well as the restricted sandwich estimator calculated under the null hypothesis. The package also includes functions for simulating data from the gamma frailty model, making custom simulations possible. For detailed help files and further code examples, we refer to the package `spass`.

An adjustment for improved type I error rate control in small sample sizes was presented. The estimation of the sandwich estimator under the null hypothesis improved type I error rate control. Controlling the type I error rate for low sample sizes is

especially important for phase II clinical trials in MS, as many of these trials tend to have low sample sizes. However, further methods for small sample sizes exist which were not compared in the scope of this work.³⁸ A comparison of these different approaches may yield interesting results.

One aspect which was not considered within this paper is the subject of unblinded sample size reestimation. When unblinding a trial, the effect size can also be estimated, and the sample size can be reestimated not only assuming the effect size is correct, but using the estimate of the effect size. An unblinded interim analysis would allow to stop for futility, making it attractive from an ethical point of view. However, the nominal testing level would have to be adjusted to avoid type I error rate inflation.

A further aspect which has received considerable interest is the incorporation of covariates, which can be included in the trend models with the addition of extra nuisance parameters. How this impacts sample size estimation and blinded sample size reestimation within the gamma frailty framework is subject to future research.

ACKNOWLEDGMENTS

This work was supported by the Deutsche Forschungsgemeinschaft (grant number FR 3070/1-1).

Conflict of interest

The authors declare no potential conflict of interests.

How to cite this article: T. Asendorf, R. Henderson, H. Schmidli, and T. Friede (2017), Sample size reestimation for clinical trials with longitudinal negative binomial counts including time trends, *Stat Med*, <vol>.

APPENDIX

A GENERAL REMARKS ON EMPLOYED DISTRIBUTIONS

A.1 Negative binomial distribution

The negative binomial distribution can be parametrised in several ways. Our form is

$$P_{NB}(x, \mu, \eta) = \left(\frac{\mu}{\mu + \eta} \right)^x \frac{\gamma(x + \eta)}{\gamma(x + 1)\gamma(\eta)} \left(\frac{\eta}{\mu + \eta} \right)^\eta,$$

with mean and variance μ and $\mu^2/\eta + \mu$, respectively.

A.2 Pairwise probabilities in gamma frailty model

Let $X_{ij}^{(t)}$ and $X_{ij}^{(s)}$ denote two observations from the gamma frailty model described in Section 3.1. Then the probability function of two realisations $x_{ij}^{(t)}$ and $x_{ij}^{(s)}$ is

$$\begin{aligned} P_{\text{NB}}^{\text{pair}} \left(x_{ij}^{(t)}, x_{ij}^{(s)}; \rho, \mu_i^{(t)}, \mu_i^{(s)}, \eta \right) &= \sum_{k=0}^{x_{ij}^{(s)}} \sum_{l=0}^{x_{ij}^{(t)}} P_{\text{NB}}(k, \mu_i^{(s)}(1 - \rho^{|s-t|}), \eta(1 - \rho^{|s-t|})) \\ &\quad \cdot P_{\text{NB}}(k, \mu_i^{(t)}(1 - \rho^{|s-t|}), \eta(1 - \rho^{|s-t|})) \\ &\quad \cdot P_{\text{NB}}(x_{ij}^{(t)} + x_{ij}^{(s)} - k - l, (\mu_i^{(s)} + \mu_i^{(t)})(1 - \rho^{|s-t|}), \eta\rho^{|s-t|}) \\ &\quad \cdot P_{\text{BIN}}(x_{ij}^{(s)} - k, x_{ij}^{(s)} + x_{ij}^{(t)} - k - l, \frac{\mu_i^{(s)}}{\mu_i^{(s)} + \mu_i^{(t)}}) \end{aligned}$$

where $P_{\text{BIN}}(k, n, p)$ denotes the probability function of a binomial distributed random variable. For further information, especially on the derivation of this formula, we refer to Fiocco et al.⁸

B CALCULATION OF THE LIKELIHOOD GRADIENT

The likelihood gradient is defined as a vector of partial derivatives:

$$\nabla_{(\eta, \lambda)} l(\eta, \lambda | \mathbf{x}) = \left(\frac{\partial l(\eta, \lambda | \mathbf{x})}{\partial \eta}, \frac{\partial l(\eta, \lambda | \mathbf{x})}{\partial \lambda_1}, \dots, \frac{\partial l(\eta, \lambda | \mathbf{x})}{\partial \lambda_D} \right)^\top$$

Therefore, calculating the gradient amounts to calculating all partial derivatives involved. This is explicitly done in the following.

$$\begin{aligned} \frac{\partial l(\eta, \lambda | \mathbf{x})}{\partial \eta} &= \frac{1}{n_C + n_E} \sum_{i=E, C} \sum_{j=1}^{n_i} \sum_{t=1}^T \frac{\partial \ln \left(P_{\text{NB}} \left(x_{ij}^{(t)}, f_i^{(t)}(\lambda), \eta \right) \right)}{\partial \eta} \\ &= \frac{1}{n_C + n_E} \sum_{i=E, C} \sum_{j=1}^{n_i} \sum_{t=1}^T 1 - \frac{x_{ij}^{(t)} + \eta}{f_i^{(t)}(\lambda) + \eta} + \ln \left(\frac{\eta}{f_i^{(t)}(\lambda) + \eta} \right) + \psi(x_{ij}^{(t)} + \eta) - \psi(\eta) \\ \frac{\partial l(\eta, \lambda | \mathbf{x})}{\partial \lambda_d} &= \frac{1}{n_C + n_E} \sum_{i=E, C} \sum_{j=1}^{n_i} \sum_{t=1}^T \frac{\partial \ln \left(P_{\text{NB}} \left(x_{ij}^{(t)}, f_i^{(t)}(\lambda), \eta \right) \right)}{\partial \lambda_d} \\ &= \frac{1}{n_C + n_E} \sum_{i=E, C} \sum_{j=1}^{n_i} \sum_{t=1}^T \frac{\partial \ln \left(P_{\text{NB}} \left(x_{ij}^{(t)}, f_i^{(t)}(\lambda), \eta \right) \right)}{\partial f_i^{(t)}(\lambda)} \cdot \frac{\partial f_i^{(t)}(\lambda)}{\partial \lambda_d} \\ &= \frac{1}{n_C + n_E} \sum_{i=E, C} \sum_{j=1}^{n_i} \sum_{t=1}^T \left(\frac{x_{ij}^{(t)}}{f_i^{(t)}(\lambda)} - \frac{x_{ij}^{(t)} + \eta}{f_i^{(t)}(\lambda) + \eta} \right) \cdot \frac{\partial f_i^{(t)}(\lambda)}{\partial \lambda_d} \end{aligned}$$

C CALCULATION OF THE LIKELIHOOD HESSIAN

The Hessian is a matrix of all second derivatives:

$$\nabla_{(\eta, \lambda)(\eta, \lambda)} l(\eta, \lambda | \mathbf{x}) = \begin{pmatrix} \frac{\partial^2 l(\eta, \lambda | \mathbf{x})}{\partial \eta^2} & \left(\frac{\partial^2 l(\eta, \lambda | \mathbf{x})}{\partial \eta \partial \lambda_d} \right)_{d=1, \dots, D}^\top \\ \left(\frac{\partial^2 l(\eta, \lambda | \mathbf{x})}{\partial \eta \partial \lambda_d} \right)_{d=1, \dots, D} & \left(\frac{\partial^2 l(\eta, \lambda | \mathbf{x})}{\partial \lambda_d \partial \lambda_e} \right)_{d=1, \dots, D, e=1, \dots, D} \end{pmatrix}.$$

The required terms follow.

$$\begin{aligned} \frac{\partial^2 l(\eta, \lambda | \mathbf{x})}{\partial \eta^2} &= \frac{1}{n_C + n_E} \sum_{i=E, C} \sum_{j=1}^{n_i} \sum_{t=1}^T \frac{\partial^2 \ln \left(P_{NB}(x_{ij}^{(t)}, f_i^{(t)}(\lambda), \eta) \right)}{\partial \eta^2} \\ &= \frac{1}{n_C + n_E} \sum_{i=E, C} \sum_{j=1}^{n_i} \sum_{t=1}^T \frac{x_{ij}^{(t)} - f_i^{(t)}(\lambda)}{(f_i^{(t)}(\lambda) + \eta)^2} + \frac{1}{\eta} - \frac{1}{f_i^{(t)}(\lambda) + \eta} + \psi_1(x_{ij}^{(t)} + \eta) - \psi_1(\eta) \\ \frac{\partial^2 l(\eta, \lambda | \mathbf{x})}{\partial \lambda_d \partial \eta} &= \frac{1}{n_C + n_E} \sum_{i=E, C} \sum_{j=1}^{n_i} \sum_{t=1}^T \frac{\partial^2 \ln \left(P_{NB}(x_{ij}^{(t)}, f_i^{(t)}(\lambda), \eta) \right)}{\partial \eta \partial f_i^{(t)}(\lambda)} \cdot \frac{f_i^{(t)}(\lambda)}{\partial \lambda_d} \\ &= \frac{1}{n_C + n_E} \sum_{i=E, C} \sum_{j=1}^{n_i} \sum_{t=1}^T \frac{x_{ij}^{(t)} - f_i^{(t)}(\lambda)}{(f_i^{(t)}(\lambda) + \eta)^2} \cdot \frac{f_i^{(t)}(\lambda)}{\partial \lambda_d} \\ \frac{\partial^2 l(\eta, \lambda | \mathbf{x})}{\partial \lambda_d \partial \lambda_e} &= \frac{1}{n_E + n_C} \sum_{i=E, C} \sum_{j=1}^{n_i} \sum_{t=1}^T \frac{\partial^2 \ln \left(P_{NB}(x_{ij}^{(t)}, f_i^{(t)}(\lambda), \eta) \right)}{\partial f_i^{(t)}(\lambda)^2} \cdot \frac{\partial f_i^{(t)}(\lambda)}{\partial \lambda_d} \cdot \frac{\partial f_i^{(t)}(\lambda)}{\partial \lambda_e} \\ &\quad + \frac{\partial \ln \left(P_{NB}(x_{ij}^{(t)}, f_i^{(t)}(\lambda), \eta) \right)}{\partial f_i^{(t)}(\lambda)} \cdot \frac{\partial^2 f_i^{(t)}(\lambda)}{\partial \lambda_d \partial \lambda_e} \\ &= \frac{1}{n_E + n_C} \sum_{i=E, C} \sum_{j=1}^{n_i} \sum_{t=1}^T \left(\frac{x_{ij}^{(t)} + \eta}{(f_i^{(t)}(\lambda) + \eta)^2} - \frac{x_{ij}^{(t)}}{f_i^{(t)}(\lambda)^2} \right) \cdot \frac{\partial f_i^{(t)}(\lambda)}{\partial \lambda_d} \cdot \frac{\partial f_i^{(t)}(\lambda)}{\partial \lambda_e} \\ &\quad + \left(\frac{x_{ij}^{(t)}}{f_i^{(t)}(\lambda)} - \frac{x_{ij}^{(t)} + \eta}{f_i^{(t)}(\lambda) + \eta} \right) \cdot \frac{\partial^2 f_i^{(t)}(\lambda)}{\partial \lambda_d \partial \lambda_e} \end{aligned}$$

D CALCULATION OF THE SANDWICH ESTIMATOR

For calculating the sandwich estimator, we require the expected Hessian matrix H as well as the expected values of the outer gradient product matrix J , which are defined as:

$$H = \frac{1}{n_C + n_E} \sum_{i=E, C} n_i \cdot E \left[\nabla_{(\eta, \lambda)(\eta, \lambda)} l_{ij}(\eta, \lambda | \mathbf{x}) \right]$$

and

$$J = \frac{1}{n_C + n_E} \sum_{i=E, C} n_i \cdot E \left[\nabla_{(\eta, \lambda)} l_{ij}(\eta, \lambda | \mathbf{x})^\top \cdot \nabla_{(\eta, \lambda)} l_{ij}(\eta, \lambda | \mathbf{x}) \right].$$

Explicit calculations of H and J are provided in the following.

D.1 Expected Hessian matrix

The expected Hessian matrix H is calculated by taking the expected values of all matrix entries from the likelihood Hessian.

These are:

$$\begin{aligned} E \left[\frac{\partial^2 l_{ij}(\eta, \lambda | \mathbf{x})}{\partial \eta^2} \right] &= \sum_{t=1}^T \frac{1}{\eta} - \frac{1}{f_i^{(t)}(\lambda) + \eta} + E[\psi_1(x_{ij}^{(t)} + \eta)] - \psi_1(\eta) \\ E \left[\frac{\partial^2 l_{ij}(\eta, \lambda | \mathbf{x})}{\partial \eta \partial \lambda_d} \right] &= 0 \\ E \left[\frac{\partial^2 l_{ij}(\eta, \lambda | \mathbf{x})}{\partial \lambda_d \partial \lambda_e} \right] &= \sum_{t=1}^T \left(\frac{1}{f_i^{(t)}(\lambda) + \eta} - \frac{1}{f_i^{(t)}(\lambda)} \right) \cdot \frac{\partial f_i^{(t)}(\lambda)}{\partial \lambda_d} \cdot \frac{\partial f_i^{(t)}(\lambda)}{\partial \lambda_e} \end{aligned}$$

D.2 Expected outer gradient matrix

The expected outer gradient matrix is calculated by taking the expected values of all matrix entries from the outer gradient product. More precisely, we would like to calculate:

$$E \left[\nabla_{(\eta, \lambda)} l_{ij}(\eta, \lambda | \mathbf{x})^\top \cdot \nabla_{(\eta, \lambda)} l_{ij}(\eta, \lambda | \mathbf{x}) \right] = E \left[\begin{pmatrix} \left(\frac{\partial l_{ij}(\eta, \lambda | \mathbf{x})}{\partial \eta} \right)^2 & \left(\frac{\partial l_{ij}(\eta, \lambda | \mathbf{x})}{\partial \eta} \cdot \frac{\partial l_{ij}(\eta, \lambda | \mathbf{x})}{\partial \lambda_d} \right)_{d=1, \dots, D}^\top \\ \left(\frac{\partial l_{ij}(\eta, \lambda | \mathbf{x})}{\partial \eta} \cdot \frac{\partial l_{ij}(\eta, \lambda | \mathbf{x})}{\partial \lambda_d} \right)_{d=1, \dots, D} & \left(\frac{\partial l_{ij}(\eta, \lambda | \mathbf{x})}{\partial \lambda_d} \cdot \frac{\partial l_{ij}(\eta, \lambda | \mathbf{x})}{\partial \lambda_e} \right)_{d=1, \dots, D}^{e=1, \dots, D} \end{pmatrix} \right]$$

This is done in the following:

$$\begin{aligned} E \left[\left(\frac{\partial l_{ij}(\eta, \lambda | \mathbf{x})}{\partial \eta} \right)^2 \right] &= E \left[\sum_{t=1}^T \sum_{s=1}^T \left(1 - \frac{x_{ij}^{(t)} + \eta}{f_i^{(t)}(\lambda) + \eta} + \ln \left(\frac{\eta}{f_i^{(t)}(\lambda) + \eta} \right) + \psi(x_{ij}^{(t)} + \eta) - \psi(\eta) \right) \right. \\ &\quad \cdot \left. \left(1 - \frac{x_{ij}^{(s)} + \eta}{f_i^{(s)}(\lambda) + \eta} + \ln \left(\frac{\eta}{f_i^{(s)}(\lambda) + \eta} \right) + \psi(x_{ij}^{(s)} + \eta) - \psi(\eta) \right) \right] \\ &= \sum_{t=1}^T \sum_{s=1}^T \left(1 + \ln \left(\frac{\eta}{f_i^{(t)}(\lambda) + \eta} \right) - \psi(\eta) \right) \cdot \left(\ln \left(\frac{\eta}{f_i^{(s)}(\lambda) + \eta} \right) - \psi(\eta) + E[\psi(x_{ij}^{(s)} + \eta)] \right) \\ &\quad - \left(1 - \frac{E[(x_{ij}^{(t)} + \eta)(x_{ij}^{(s)} + \eta)]}{(f_i^{(t)}(\lambda) + \eta)(f_i^{(s)}(\lambda) + \eta)} + \ln \left(\frac{\eta}{f_i^{(s)}(\lambda) + \eta} \right) - \psi(\eta) + \frac{E[(x_{ij}^{(t)} + \eta)\psi(x_{ij}^{(s)} + \eta)]}{f_i^{(t)}(\lambda) + \eta} \right) \\ &\quad + E[\psi(x_{ij}^{(t)} + \eta)] - \frac{E[\psi(x_{ij}^{(t)} + \eta)(x_{ij}^{(s)} + \eta)]}{f_i^{(s)}(\lambda) + \eta} + E[\psi(x_{ij}^{(t)} + \eta)] \cdot \left(\ln \left(\frac{\eta}{f_i^{(s)}(\lambda) + \eta} \right) - \psi(\eta) \right) \\ &\quad + E[\psi(x_{ij}^{(t)} + \eta)\psi(x_{ij}^{(s)} + \eta)] \\ &= \sum_{t=1}^T \sum_{s=1}^T \left(1 + \ln \left(\frac{\eta}{f_i^{(t)}(\lambda) + \eta} \right) - \psi(\eta) \right) \cdot \left(\ln \left(\frac{\eta}{f_i^{(s)}(\lambda) + \eta} \right) - \psi(\eta) + E[\psi(x_{ij}^{(s)} + \eta)] \right) \end{aligned}$$

$$\begin{aligned}
& + (1 + \ln\left(\frac{\eta}{f_i^{(s)}(\lambda) + \eta}\right) - \psi(\eta)) \cdot (E[\psi(x_{ij}^{(t)} + \eta)] - 1) \\
& + \frac{E[(x_{ij}^{(t)} x_{ij}^{(s)}) + \eta^2 + \eta f_i^{(t)}(\lambda) + \eta f_i^{(s)}(\lambda)]}{(f_i^{(t)}(\lambda) + \eta)(f_i^{(s)}(\lambda) + \eta)} \\
& - \frac{E[(x_{ij}^{(t)} + \eta)\psi(x_{ij}^{(s)} + \eta)]}{f_i^{(t)}(\lambda) + \eta} - \frac{E[\psi(x_{ij}^{(t)} + \eta)(x_{ij}^{(s)} + \eta)]}{f_i^{(s)}(\lambda) + \eta} \\
& + E[\psi(x_{ij}^{(t)} + \eta)\psi(x_{ij}^{(s)} + \eta)] \\
& = \sum_{t=1}^T \sum_{s=1; s \neq t}^T E[\psi(x_{ij}^{(s)} + \eta)] \cdot \left(1 + \ln\left(\frac{\eta}{f_i^{(t)}(\lambda) + \eta}\right) - \psi(\eta) - \frac{\eta}{f_i^{(t)}(\lambda) + \eta}\right) \\
& + E[\psi(x_{ij}^{(t)} + \eta)] \cdot \left(1 + \ln\left(\frac{\eta}{f_i^{(s)}(\lambda) + \eta}\right) - \psi(\eta) - \frac{\eta}{f_i^{(s)}(\lambda) + \eta}\right) \\
& - \frac{E[x_{ij}^{(t)}\psi(x_{ij}^{(s)} + \eta)]}{f_i^{(t)}(\lambda) + \eta} - \frac{E[x_{ij}^{(s)}\psi(x_{ij}^{(t)} + \eta)]}{f_i^{(s)}(\lambda) + \eta} + E[\psi(x_{ij}^{(s)} + \eta)\psi(x_{ij}^{(t)} + \eta)] \\
& + \left(\left(\frac{\eta}{f_i^{(t)}(\lambda) + \eta}\right) - \psi(\eta)\right) \left(\left(\frac{\eta}{f_i^{(s)}(\lambda) + \eta}\right) - \psi(\eta)\right) + \frac{\rho^{|s-t|} f_i^{(t)}(\lambda) f_i^{(s)}(\lambda)}{\eta(f_i^{(t)}(\lambda) + \eta)(f_i^{(s)}(\lambda) + \eta)} \\
& + \sum_{t=1}^T 2 \cdot E[\psi(x_{ij}^{(t)} + \eta)] \cdot \left(1 + \ln\left(\frac{\eta}{f_i^{(t)}(\lambda) + \eta}\right) - \psi(\eta) - \frac{\eta}{f_i^{(t)}(\lambda) + \eta}\right) \\
& - 2 \cdot \frac{E[x_{ij}^{(t)}\psi(x_{ij}^{(t)} + \eta)]}{f_i^{(t)}(\lambda) + \eta} + E[\psi(x_{ij}^{(t)} + \eta)^2] \\
& + \left(\left(\frac{\eta}{f_i^{(t)}(\lambda) + \eta}\right) - \psi(\eta)\right)^2 + \frac{f_i^{(t)}(\lambda)}{\eta(f_i^{(t)}(\lambda) + \eta)} \\
E\left[\frac{\partial l_{ij}(\eta, \lambda | \mathbf{x})}{\partial \eta} \cdot \frac{\partial l_{ij}(\eta, \lambda | \mathbf{x})}{\partial \lambda_d}\right] & = \sum_{t=1}^T \sum_{s=1}^T E\left[\left(1 - \frac{x_{ij}^{(t)} + \eta}{f_i^{(t)}(\lambda) + \eta} + \ln\left(\frac{\eta}{f_i^{(t)}(\lambda) + \eta}\right) + \psi(x_{ij}^{(t)} + \eta) - \psi(\eta)\right)\right. \\
& \cdot \left.\left(\frac{x_{ij}^{(s)}}{f_i^{(s)}(\lambda)} - \frac{x_{ij}^{(s)} + \eta}{f_i^{(s)}(\lambda) + \eta}\right) \cdot \frac{\partial f_i^{(s)}(\lambda)}{\partial \lambda_d}\right] \\
& = \sum_{t=1}^T \sum_{s=1}^T E\left[\left(-\frac{x_{ij}^{(t)} + \eta}{f_i^{(t)}(\lambda) + \eta} + \psi(x_{ij}^{(t)} + \eta)\right) \cdot \left(\frac{x_{ij}^{(s)}}{f_i^{(s)}(\lambda)} - \frac{x_{ij}^{(s)} + \eta}{f_i^{(s)}(\lambda) + \eta}\right) \cdot \frac{\partial f_i^{(s)}(\lambda)}{\partial \lambda_d}\right] \\
& = \sum_{t=1}^T \sum_{s=1}^T \frac{E[-(x_{ij}^{(t)} + \eta)x_{ij}^{(s)}(f_i^{(s)}(\lambda) + \eta) + (x_{ij}^{(t)} + \eta)(x_{ij}^{(s)} + \eta)f_i^{(s)}(\lambda)]}{(f_i^{(t)}(\lambda) + \eta)(f_i^{(s)}(\lambda) + \eta)f_i^{(s)}(\lambda)} \cdot \frac{\partial f_i^{(s)}(\lambda)}{\partial \lambda_d} \\
& + E[\psi(x_{ij}^{(t)} + \eta)x_{ij}^{(s)}] \left(\frac{1}{f_i^{(s)}(\lambda)} - \frac{1}{f_i^{(s)}(\lambda) + \eta}\right) \cdot \frac{\partial f_i^{(s)}(\lambda)}{\partial \lambda_d} \\
& - E[\psi(x_{ij}^{(s)} + \eta)] \cdot \frac{\eta}{f_i^{(s)}(\lambda) + \eta} \cdot \frac{\partial f_i^{(s)}(\lambda)}{\partial \lambda_d} \\
& = \sum_{t=1}^T \sum_{s=1; s \neq t}^T \frac{-\rho^{|t-s|} f_i^{(t)}(\lambda)}{(f_i^{(t)}(\lambda) + \eta)(f_i^{(s)}(\lambda) + \eta)} \cdot \frac{\partial f_i^{(s)}(\lambda)}{\partial \lambda_d} \\
& + \left(E[\psi(x_{ij}^{(t)} + \eta)x_{ij}^{(s)}] \cdot \frac{\eta}{f_i^{(s)}(\lambda)(f_i^{(s)}(\lambda) + \eta)} - E[\psi(x_{ij}^{(t)} + \eta)] \frac{\eta}{f_i^{(s)}(\lambda) + \eta}\right) \cdot \frac{\partial f_i^{(s)}(\lambda)}{\partial \lambda_d}
\end{aligned}$$

$$\begin{aligned}
& + \sum_{t=1}^T -\frac{1}{(f_i^{(t)}(\lambda) + \eta)} \cdot \frac{\partial f_i^{(t)}(\lambda)}{\partial \lambda_d} \\
& + \left(E[\psi(x_{ij}^{(t)} + \eta)x_{ij}^{(t)}] \cdot \frac{\eta}{f_i^{(t)}(\lambda)(f_i^{(t)}(\lambda) + \eta)} - E[\psi(x_{ij}^{(t)} + \eta)] \frac{\eta}{f_i^{(t)}(\lambda) + \eta} \right) \cdot \frac{\partial f_i^{(t)}(\lambda)}{\partial \lambda_d} \\
E\left[\frac{\partial l_{ij}(\eta, \lambda|\mathbf{x})}{\partial \lambda_d} \cdot \frac{\partial l_{ij}(\eta, \lambda|\mathbf{x})}{\partial \lambda_e}\right] & = \sum_{t=1}^T \sum_{s=1}^T E\left[\left(\frac{x_{ij}^{(t)}}{f_i^{(t)}(\lambda)} - \frac{x_{ij}^{(t)} + \eta}{f_i^{(t)}(\lambda) + \eta}\right) \cdot \frac{\partial f_i^{(t)}(\lambda)}{\partial \lambda_d}\right. \\
& \quad \cdot \left.\left(\frac{x_{ij}^{(s)}}{f_i^{(s)}(\lambda)} - \frac{x_{ij}^{(s)} + \eta}{f_i^{(s)}(\lambda) + \eta}\right) \cdot \frac{\partial f_i^{(s)}(\lambda)}{\partial \lambda_e}\right] \\
& = \sum_{t=1}^T \sum_{s=1; s \neq t}^T \rho^{|t-s|} \frac{\eta}{(f_i^{(t)}(\lambda) + \eta)(f_i^{(s)}(\lambda) + \eta)} \cdot \frac{\partial f_i^{(t)}(\lambda)}{\partial \lambda_d} \cdot \frac{\partial f_i^{(s)}(\lambda)}{\partial \lambda_e} \\
& + \sum_{t=1}^T \frac{\eta}{f_i^{(t)}(\lambda)(f_i^{(t)}(\lambda) + \eta)} \cdot \frac{\partial f_i^{(t)}(\lambda)}{\partial \lambda_d} \cdot \frac{\partial f_i^{(t)}(\lambda)}{\partial \lambda_e}
\end{aligned}$$

E ASYMPTOTIC PROPERTIES OF PL ESTIMATORS

In the following we derive asymptotic properties for the presented likelihood, following standard likelihood procedures. Let $\hat{\eta}$ and $\hat{\lambda}$ denote the pseudolikelihood estimates of η and λ respectively, i.e.:

$$(\hat{\eta}, \hat{\lambda}) = \underset{(\eta, \lambda) \in \mathbb{R}_+^{D+1}}{\operatorname{argmin}} l(\eta, \lambda|\mathbf{x}).$$

Then it immediately follows that $\nabla_{(\eta, \lambda)} l(\hat{\eta}, \hat{\lambda}|\mathbf{x}) = 0$. Using this property and the multivariate mean value theorem, we derive the following formula:

$$\begin{aligned}
& \nabla_{(\eta, \lambda)(\eta, \lambda)} l(\tilde{\eta}, \tilde{\lambda}|\mathbf{x}) \cdot ((\hat{\eta}, \hat{\lambda}) - (\eta, \lambda)) = \nabla_{(\eta, \lambda)} l(\hat{\eta}, \hat{\lambda}|\mathbf{x}) - \nabla_{(\eta, \lambda)} l(\eta, \lambda|\mathbf{x}) \\
\Leftrightarrow & \quad ((\hat{\eta}, \hat{\lambda}) - (\eta, \lambda)) = - [\nabla_{(\eta, \lambda)(\eta, \lambda)} l(\tilde{\eta}, \tilde{\lambda}|\mathbf{x})]^{-1} \cdot \nabla_{(\eta, \lambda)} l(\eta, \lambda|\mathbf{x}) \\
\Leftrightarrow & \quad \sqrt{n_C + n_E} \cdot ((\hat{\eta}, \hat{\lambda}) - (\eta, \lambda)) = - [\nabla_{(\eta, \lambda)(\eta, \lambda)} l(\tilde{\eta}, \tilde{\lambda}|\mathbf{x})]^{-1} \cdot \sqrt{n_C + n_E} \cdot \nabla_{(\eta, \lambda)} l(\eta, \lambda|\mathbf{x})
\end{aligned}$$

where $(\tilde{\eta}, \tilde{\lambda})$ denotes some point between $(\hat{\eta}, \hat{\lambda})$ and (η, λ) which fulfills the equation. The existence of such a point is guaranteed by the multivariate mean value theorem. For convenience, denote:

$$l_{ij}(\eta, \lambda|\mathbf{x}) = \sum_{t=1}^T \ln \left(P_{\text{NB}} \left(x_{ij}^{(t)}, f_i^{(t)}(\lambda), \eta \right) \right).$$

Then we can write the equation as:

$$\Leftrightarrow \sqrt{n_C + n_E} ((\hat{\eta}, \hat{\lambda}) - (\eta, \lambda)) = - \underbrace{\left[\frac{1}{n_C + n_E} \sum_{i=E, C} \sum_{j=1}^{n_i} \nabla_{(\eta, \lambda)(\eta, \lambda)} l_{ij}(\tilde{\eta}, \tilde{\lambda}|\mathbf{x}) \right]^{-1}}_1$$

$$\cdot \underbrace{\frac{1}{\sqrt{n_C + n_E}} \sum_{i=E,C} \sum_{j=1}^{n_i} \nabla_{(\eta,\lambda)} l_{ij}(\eta, \lambda | \mathbf{x})}_2$$

Assuming a constant sample size allocation factor $n_E/n_C = k$, by the law of large numbers the first term converges to:

$$\begin{aligned} \frac{1}{n_C + n_E} \sum_{i=E,C} \sum_{j=1}^{n_i} \nabla_{(\eta,\lambda)(\eta,\lambda)} l_{ij}(\tilde{\eta}, \tilde{\lambda} | \mathbf{x}) \xrightarrow{n_E, n_C \rightarrow \infty} \frac{k}{1+k} E [\nabla_{(\eta,\lambda)(\eta,\lambda)} l_{Ej}(\eta, \lambda | \mathbf{x})] \\ + \frac{1}{1+k} E [\nabla_{(\eta,\lambda)(\eta,\lambda)} l_{Cj}(\eta, \lambda | \mathbf{x})] = H \end{aligned}$$

For the second term it holds by the central limit theorem that:

$$\begin{aligned} \frac{1}{\sqrt{n_C + n_E}} \sum_{i=E,C} \sum_{j=1}^{n_i} \nabla_{(\eta,\lambda)} l_{ij}(\eta, \lambda | \mathbf{x}) \rightarrow N(0, \frac{k}{1+k} E [\nabla_{(\eta,\lambda)} l_{Ej}(\eta, \lambda | \mathbf{x})^\top \cdot \nabla_{(\eta,\lambda)} l_{Ej}(\eta, \lambda | \mathbf{x})] \\ + \frac{1}{1+k} E [\nabla_{(\eta,\lambda)} l_{Cj}(\eta, \lambda | \mathbf{x})^\top \cdot \nabla_{(\eta,\lambda)} l_{Cj}(\eta, \lambda | \mathbf{x})]) = N(0, J) \end{aligned}$$

By Slutsky's Theorem we then attain that:

$$\sqrt{n_C + n_E} \cdot ((\hat{\eta}, \hat{\lambda}) - (\eta, \lambda)) \rightarrow N(0, H^{-1} J H^{-1})$$

Where the matrix H is the expected Hessian of the likelihood, J is the expected outer gradient matrix and $H^{-1} J H^{-1}$ is the sandwich estimator.

F APPLICATION OF THE PROCEDURE IN R

In this section we give a detailed demonstration of the sample size estimation and reestimation techniques available for the gamma frailty model in the R-package `spass`. Specifically for the gamma frailty model described in this publication, the R-package `spass` contains multiple functions:

- `rnbinom.gf()` for simulating data from the described gamma frailty model
- `fit.nb.gf()` for fitting data with a gamma frailty model
- `n.nb.gf()` for initial sample size estimation
- `bssr.nb.gf()` for blinded sample size reestimation
- `test.nb.gf()` for testing the null hypothesis within the gamma frailty model

We now demonstrate the full procedure of sample size estimation and blinded sample size reestimation using the same data from Fernandez et al.³³ as in the application from Section 5.4, i.e. comparing the low-dose treatment group and the placebo group.

F.1 Initial sample size estimation

Let us assume that an MS trial is being planned and initial data, in this case consisting of the placebo and low-dose group from Fernandez et al.,³³ has been collected in a pilot study. To assess the required parameters for sample size estimation, assuming an underlying exponential trend, the data can be fitted using the function `fit.nb.gf()` as follows:

```
fit.nb.gf(dataC=dataC, dataE=dataE, trend="exponential", rho=T)
```

```
$par
```

```
      lambda1      lambda2      lambda3      size
-0.1915827  0.2616038 -0.3927572  0.4953651
```

```
$value
```

```
[1] 3.87595
```

```
$counts
```

```
function gradient
```

```
      28      28
```

```
$convergence
```

```
[1] 0
```

```
$message
```

```
[1] "CONVERGENCE: REL_REDUCTION_OF_F <= FACTR*EPSMCH"
```

```
$hessian
```

```
      [,1]      [,2]      [,3]      [,4]
[1,] -0.9512230686 -0.964165621 -0.4055958993  0.0001999866
[2,] -0.9641656211 -1.579388302 -0.6624096683 -0.0026141294
[3,] -0.4055958993 -0.662409668 -0.6624096683  0.0009456138
[4,]  0.0001999866 -0.002614129  0.0009456138 -1.2254941300
```

```
$ogradient
      [,1]      [,2]      [,3]      [,4]
[1,] 1.5503759 1.5667449 0.7766687 1.1116493
[2,] 1.5667449 1.7490272 0.7601481 0.9890489
[3,] 0.7766687 0.7601481 0.7601481 0.1930255
[4,] 1.1116493 0.9890489 0.1930255 2.1050299
```

```
$rho
```

```
[1] 1
```

The function takes the data in wide format for each group, the control group (dataC) and the experiment group (dataE) separately. Further, the intended trend is specified and an estimation of the correlation parameter ρ is demanded. The result reveals fitted values of $\lambda_1, \dots, \lambda_3$ and the shape parameter η , denoted as size. The correlation parameter is given under rho. In this case, the correlation parameter is estimated to be 1, the highest possible estimate in the gamma frailty model, indicating a constant frailty term as appropriate.

In a second step, the results from the fitted model are passed on to the function `n.nb.gf()` for sample size estimation. Assuming the rates and shape parameter will hold in a planned larger trial with $T = 3$ time points, equal group sizes ($k = 1$) and a slightly larger effect, we would like to reject the null hypothesis $H_0 : h(\eta, \lambda) = \lambda_3 \geq 0 = h_0$ at level $\alpha = 2.5\%$ (one-sided) with power of 80%. The required sample size is calculated as follows:

```
h <- function(lambda.eta){lambda.eta[3]}
hgrad <- function(lambda.eta){c(0,0,1,0)}
ssEst <- n.nb.gf(alpha=0.025, power=0.8, lambda=c(-0.20, 0.25, -0.70),
               size=0.50, rho=1, tp=3, k=1, h=h,
               hgrad=hgrad, h0=0, trend="exponential")
summary(ssEst)
```

```
Initial Sample Size Calculation
```

```
-----
alpha level:           0.025
testing power:         0.8
trend type:            exponential
```

```

trend parameters:      -0.2 0.25 -0.7
dispersion parameter:  0.5
correlation parameter: 1
time points:           3
allocation factor:      1

```

Sample Size

```

-----
control group:    34.91
treatment group: 34.91

```

The output reveals that a study with these assumptions would require at least 35 patients per group to prove the clinically relevant effect size with 80% power. Note that prior to performing the sample size estimation, the functions `h()` and its gradient `hgrad()` were specified. The function `h` is a function in which the first arguments always refer to λ and the last argument refers to the shape parameter η . In this case, we are only interested in testing $H_0 : \lambda_3 \geq 0$, therefore `h()` is defined in such a way that the third entry is returned. For further details on the required arguments and output of the function `n.nb.gf()` we refer to the help files of the R-package `spass`.

F.2 Blinded sample size reestimation

For demonstrating the blinded sample size reestimation we pool the data from the two groups and consider this to be the blinded data on which a blinded sample size reestimation is to be performed. The data given to the statistician, now without information on the group assignment, is passed to the function `bssr.nb.gf()` and yields following result:

```

bssrEst <- bssr.nb.gf(data=daten.bssr, alpha=0.025, power=0.8, delta=-0.7, h0=0,
                      tp=3, k=1, trend="exponential")
summary(bssrEst)

```

Blinded Sample Size Reestimation

```

-----
alpha level:           0.025
testing power:          0.8
trend type:             exponential

```

```
effect size: -0.7
est. trend parameters: -0.18 0.36 -0.7
est. dispersion parameter: 0.54
est. correlation parameter: 1
time points: 3
allocation factor: 1
```

Sample Size

```
-----
control group: 30.82
treatment group: 30.82
```

Blinded Sample Size Reestimation

```
-----
alpha level: 0.025
testing power: 0.8
trend type: exponential
effect size: -0.7
est. trend parameters: -0.18 0.36 -0.7
est. dispersion parameter: 0.54
est. correlation parameter: 1
time points: 3
allocation factor: 1
```

Sample Size

```
-----
control group: 30.82
treatment group: 30.82
```

A blinded sample size reestimation with the 21 enrolled patients would have revealed, that 31 patients per group would be necessary to show a significant effect size of $\lambda_3 = -0.70$ with power of 80%. The sample size would have been sufficiently large and no increase necessary.

F.3 Evaluation

At the end of the clinical trial, when the groups are unblinded, the function `test.nb.gf()` can be used for testing treatment effects. Again we refer to the data set from Fernandez et al.³³ and test for treatment differences between the placebo and the low dose strata. Then the null hypothesis $H_0 : \lambda_3 \geq 0$ can be tested as follows:

```
h <- function(lambda.eta){lambda.eta[3]}
hgrad <- function(lambda.eta){c(0,0,1,0)}
test.nb.gf(dataC=dataC, dataE=dataE, h=h, hgrad=hgrad, h0=0,
           trend="exponential", H0=TRUE, one.sided=TRUE)
```

effect	stderr	Z	p-value
-0.3927572	0.4277850	-0.9181182	0.1792785

giving us the result, that the slope of the treatment group in the exponential model is not significantly lower than the slope of the control group. The argument `H0=TRUE` was set to calculate a standard error restricted to the null hypothesis, as described in Section 7. The restriction should be made in this case, as the group sizes are relatively low and an inflation of the type I error can not be ruled out otherwise. For more information on the described functions and their arguments we refer to the help files of the R-package `spass`.

References

- [1] Guideline on clinical investigation of medicinal products for the treatment of Multiple Sclerosis. http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2015/03/WC500185161.pdf. Published March 2015. Accessed July 12, 2018.
- [2] Chataway J, Nicholas R, Todd S, et al. A novel adaptive design strategy increases the efficiency of clinical trials in secondary progressive multiple sclerosis. *Mult Scler*. 2011;17(1):81–88.
- [3] Cox DR, Miller HD. *The theory of stochastic processes*. London: Methuen and Co. Ltd.; 1965.

- [4] Jacobs P, Lewis P. Discrete time series generated by mixtures I: Correlational and runs properties. *J R Stat Soc Series B Stat Methodol.* 1978;40(1):94–105.
- [5] Jacobs PA, Lewis PAW. Discrete time series generated by mixtures II: Asymptotic properties. *J R Stat Soc Series B Stat Methodol.* 1978;40(2):222–228.
- [6] Henderson R, Shimakura S. A serially correlated gamma frailty model for longitudinal count data. *Biometrika.* 2003;90(2):355–366.
- [7] Hall P, Mueller H, Yao F. Modelling sparse generalised longitudinal observations with latent Gaussian processes. *J R Stat Soc Series B Stat Methodol.* 2008;70(4):703–723.
- [8] Fiocco M, Putter H, Van Houwelingen JC. A new serially correlated gamma-frailty process for longitudinal count data. *Biostatistics.* 2009;10(2):245–57.
- [9] Altman R, Petkau A. Application of hidden Markov models to multiple sclerosis lesion count data. *Stat Med.* 2005;24(15):2335–2344.
- [10] Altman R, Petkau A, Vrecko D, Smith A. MRI-based clinical trials in relapsing-remitting MS: new sample size calculations based on a longitudinal model. *Mult Scler.* 2012;18(11):1600–1608.
- [11] McKenzie E. Autoregressive moving-average processes with negative-binomial and geometric marginal distributions. *Adv Appl Probab.* 1986;18:679–705.
- [12] Thall PF, Vail SC. Some Covariance Models for Longitudinal Count Data with Overdispersion. *Biometrics.* 1990;46(3):657–671.
- [13] Schneider S, Schmidli H, Friede T. Blinded sample size re-estimation for recurrent event data with time trends. *Stat Med.* 2013;32(30):5448–5457.
- [14] Wittes J, Brittain E. The role of internal pilot studies in increasing the efficiency of clinical trials. *Stat Med.* 1990;9(1-2):65–72.
- [15] Proschan MA. Two-stage sample size re-estimation based on a nuisance parameter: a review. *J Biopharm Stat.* 2005;15(4):559–574.
- [16] Proschan MA. Sample size re-estimation in clinical trials. *Biom J.* 2009;51(2):348–357.
- [17] Friede T, Kieser M. Sample size recalculation in internal pilot study designs: a review. *Biom J.* 2006;48(4):537–555.

- [18] Friede T, Schmidli H. Blinded sample size reestimation with count data: methods and applications in multiple sclerosis. *Stat Med*. 2010;29(10):1145–1156.
- [19] Friede T, Schmidli H. Blinded sample size reestimation with negative binomial counts in superiority and non-inferiority trials. *Methods Inf Med*. 2010;49(6):618–624.
- [20] Asendorf T, Henderson R, Schmidli H, Friede T. Modelling and sample size reestimation for longitudinal count data with incomplete follow up. *Stat Methods Med Res*. 0;0(0):1-17. <http://dx.doi.org/10.1177/0962280217715664>. Accessed July 12, 2018.
- [21] Steinvorth SM, Rover C, Schneider S, Nicholas R, Straube S, Friede T. Explaining temporal trends in annualised relapse rates in placebo groups of randomised controlled trials in relapsing multiple sclerosis: systematic review and meta-regression. *Mult Scler*. 2013;19(12):1580–1586.
- [22] Nicholas R, Straube S, Schmidli H, Pfeiffer S, Friede T. Time-patterns of annualized relapse rates in randomized placebo-controlled clinical trials in relapsing multiple sclerosis: a systematic review and meta-analysis. *Mult Scler*. 2012;18(9):1290–1296.
- [23] Stellmann JP, Stürner KH, Young KL, Siemonsen S, Friede T, Heesen C. Regression to the mean and predictors of MRI disease activity in RRMS placebo cohorts - Is there a place for baseline-to-treatment studies in MS?. *PLoS One*. 2015;10(2):1-14.
- [24] Sormani MP, Stubinski B, Cornelisse P, Rocak S, Li D, De Stefano N. Magnetic resonance active lesions as individual-level surrogate for relapses in multiple sclerosis. *Mult Scler*. 2011;17(5):541–549.
- [25] van Noort JM, Bsibsi M, Nacken PJ, Verbeek R, Venneker EH. Therapeutic intervention in multiple sclerosis with Alpha B-Crystallin: A randomized controlled phase IIa trial. *PLoS One*. 2015;10(11):e0143366. <https://doi.org/10.1371/journal.pone.0143366>. Accessed July 12, 2018.
- [26] Kappos L, Antel J, Comi G, et al. Oral Fingolimod (FTY720) for Relapsing Multiple Sclerosis. *N Engl J Med*. 2006;355(11):1124-1140.
- [27] Comi G, Pulizzi A, Rovaris M, et al. Effect of laquinimod on MRI-monitored disease activity in patients with relapsing-remitting multiple sclerosis: a multicentre, randomised, double-blind, placebo-controlled phase IIb study. *Lancet*. 2008;371(9630):2085 - 2092.
- [28] Miller DH. Guidelines for MRI monitoring of the treatment of multiple sclerosis: recommendations of the US Multiple Sclerosis Society's task force. *Mult Scler*. 1996;1(6):335-338.

- [29] Schimrigk S, Brune N, Hellwig K, et al. Oral fumaric acid esters for the treatment of active multiple sclerosis: an open-label, baseline-controlled pilot study. *Eur J Neurol*. 2006;13(6):604–610.
- [30] Chalavi S, Simmons A, Dijkstra H, Barker GJ, Reinders AA. Quantitative and qualitative assessment of structural magnetic resonance imaging data in a two-center study. *BMC Med Imaging*. 2012;12(1):27.
- [31] Guidance for industry-adaptive design clinical trials for drugs and biologics. <http://www.fda.gov/downloads/Drugs/.../Guidances/ucm201790.pdf>. Published February 2010. Accessed July 12, 2018.
- [32] Schneider S, Schmidli H, Friede T. Robustness of methods for blinded sample size re-estimation with overdispersed count data. *Stat Med*. 2013;32(21):3623–3635.
- [33] Fernández O, Izquierdo G, Fernández V, et al. Adipose-derived mesenchymal stem cells (AdMSC) for the treatment of secondary-progressive multiple sclerosis: A triple blinded, placebo controlled, randomized phase I/II safety and feasibility study. *PLoS One*. 2018;13(5):1-14.
- [34] Friede T, Kieser M. Blinded sample size reassessment in non-inferiority and equivalence trials. *Stat Med*. 2003;22(6):995-1007.
- [35] Aban IB, Cutter GR, Mavinga N. Inferences and power analysis concerning two negative binomial distributions with an application to MRI lesion counts data. *Comput Stat Data Anal*. 2008;53(3):820–833.
- [36] Zhu H, Lakkis H. Sample size calculation for comparing two negative binomial rates. *Stat Med*. 2014;33(3):376–387.
- [37] Birkett MA, Day SJ. Internal pilot studies for estimating sample size. *Stat Med*. 1994;13(23-24):2455–2463.
- [38] Mancl LA, DeRouen TA. A covariance estimator for GEE with improved small-sample properties. *Biometrics*. 2001;57(1):126–134.
- [39] Mütze T, Munk A, Friede T. Design and analysis of three-arm trials with negative binomially distributed endpoints. *Stat Med*. 2016;35(4):505–521.
- [40] Mielke M, Munk A, Schacht A. The assessment of non-inferiority in a gold standard design with censored, exponentially distributed endpoints. *Stat Med*. 2008;27(25):5093-5110.
- [41] Wachtlin D, Kieser M. Blinded Sample Size Recalculation in Longitudinal Clinical Trials Using Generalized Estimating Equations. *Therapeutic Innovation & Regulatory Science*. 2013;47(4):460-467. PMID: 30235527.
- [42] Asendorf T, Gera R, Islam S, Harden M, Placzek M. R-package spass: study planning and adaptation of sample size. <https://CRAN.R-project.org/package=spass>. Published July 2018. Accessed July 12, 2018.